VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV INTELIGENTNÍCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF INTELLIGENT SYSTEMS

MULTIPLE-CONTEXT TRUST MODEL FOR SOCIAL NETWORKS USING PERSONALITY ANALYSIS

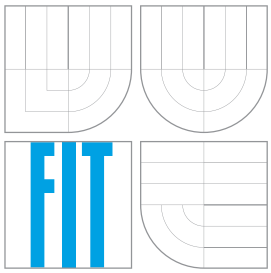
DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

Bc. TOMÁŠ ŠVEC

BRNO 2014

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV INTELIGENTNÍCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF INTELLIGENT SYSTEMS

VÍCE-KONTEXTOVÝ MODEL DŮVĚRY V SOCIÁLNÍ SÍTI S VYUŽITÍM ANALÝZY OSOBNOSTI

MULTIPLE-CONTEXT TRUST MODEL FOR SOCIAL NETWORKS USING PERSONALITY ANALYSIS

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

Bc. TOMÁŠ ŠVEC

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. JAN SAMEK, Ph.D.

BRNO 2014

Abstrakt

Tato diplomová práce navazuje na bakalářskou práci, ve které byl vytvořen model důvěry pro sociální síť Facebook. Do tohoto modelu jsou zapracovány připomínky z konference UMAP 2013 a ověřena jeho škálovatelnost a flexibilita. V další části práce jsou uvedeny základní termíny z psychologie osobnosti a zkoumána závislost důvěry na osobnosti uživatele. Je vybrán model Big Five k reprezentaci charakteru uživatele a navržen dotazník, u nějž bude zkoumána korelace s modelem důvěry. Tato korelace je na základě sociologických poznatků odhadnuta a později ověřena na reálných uživatelích sociální sítě Facebook.

Abstract

This master's thesis follows up on the bachelor thesis which described a model of trust for the social network Facebook. The model has been enhanced with remarks from the UMAP 2013 conference and its scalability and flexibility were verified. Basic terminology from personality psychology is explained and then used to find correlation between trust and users' personality. The model Big Five is chosen to represent user's character and a questionnaire is designed. Said questionnaire is used to find correlation between Big Five and the multi-context trust model. This correlation is estimated based on knowledge from the field of sociology and then verified on real Facebook users.

Klíčová slova

Důvěra, reputace, sociální síť, analýza sociální sítě, kontext důvěry, Facebook, osobnostní psychologie, Big Five, charakter.

Keywords

Trust, reputation, social network, social network analysis, context of trust, Facebook, personality psychology, Big Five, character.

Citace

Tomáš Švec: Multiple-context trust model for social networks using personality analysis, diplomová práce, Brno, FIT VUT v Brně, 2014

Multiple-context trust model for social networks using personality analysis

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně pod vedením pana Ing. Jana Samka, Ph.D.

.....
Tomáš Švec
May 27, 2014

Poděkování

Na tomto místě bych chtěl poděkovat vedoucímu mé práce, Ing. Janu Samkovi, Ph.D., za odborné vedení práce a motivaci v práci na předloženém tématu pokračovat i v magisterském studiu. Rád bych také poděkoval vedení Fakulty informačních technologií VUT v Brně za možnost téma prezentovat na konferenci UMAP 2013.

Contents

Chapter 1

Introduction

Never trust he who trusts everyone.

Carlos Ruiz Zafón, The Shadow of the Wind

In the closely networked world we live in today, there is no way of assuring the direct security mechanisms present in interpersonal communication in the history of mankind. Social networks are seizing opportunities to penetrate our lives and are also considered new tools for controlling the opinions of the public.

The faith in this type of media and internet facilities in general has been shattered lately with the leak of information from the NSA agency. When Edward Snowden first disclosed sensitive documents to journalist Glenn Greenwald in late 2012, the public confidence in online media began to deteriorate. This is not an ordinary sort of trust between individual people, this comprises the trust held by a large group of people in technology itself.

One fact the general public had been tending to forget in the past few months is that social networks are merely a tool for communication and it is solely up to the user how much information he or she discloses to the service provider. The NSA leak did show the danger of providing too much information on oneself online.

On the other hand, the provision of such information is (still) not mandatory. To conclude the idea introduced here in the quotation from Carlos Ruiz Zafón, people who trust their social network excessively are not likely to be very adept at keeping the information to themselves and therefore it is not wise to trust them with it either.

The provided information is indeed used for the inference of new knowledge. This master's thesis continues the effort of the bachelor thesis from the same author [Šve11] presented at the UMAP 2013 conference to build and validate a model of trust among users in social networks. In addition to adapting the model to the current situation and remarks from the conference, the key contribution is the addition of personality psychology into the whole process. In correlation with the findings of the aforementioned thesis, the built trust model performs differently for different personalities.

The key goal is to renovate the multi-context trust model, then explore the field of personality psychology. The establishment shall be followed by a survey of current models used to express personality traits and a subsequent selection of a suitable model for this application. The selected model shall then be implemented in the form of an online questionnaire and its results compared and correlated with the results of the given trust model on real user data.

Chapter 2

Trust, reputation and social networks

This theoretical chapter of the thesis is aimed at establishing key terms in the field of social network analysis, trust and reputation in the computational sense of the concept. The area of trust and reputation is located on the verge of two fields, sociology and artificial intelligence, therefore it is not always possible to stay within the boundaries of accurate description and mathematical definitions. To this date has never been found an apparatus to describe human emotions deterministically.

Sections are built upon dual parallelism between sociology and information technologies. The transition of trust and reputation into the digital world brought important consequences, as did the transition of social networks from physical communities into the computer network. It is this parallelism that promises results in connecting trust and social networks.

Most definitions and explanations in this chapter are taken from the bachelor thesis [Šve11].

2.1 Trust

Trust is a long-discussed concept in sociology. Bruce Schneier [Sch12] considers the ability of building trust with non-related individuals to be the cornerstone of modern society. Evolutionary biology describes a mechanism which can be traced to the ancestors of current animal species called *reciprocal altruism*. This concept is often explained on groups of animals (herds, flocks, regional units) sharing food in case one of them was unsuccessful when providing for itself. This concept seems to contradict Darwin's theory of evolution. The fact holds true for the momentary situation, but the animal anticipates similar behaviour in case the odds change and it finds itself in a similar situation.

Analyzing the term *trust* from the point of view of individual branches of science would not be wise, neither is it the goal of this thesis. The definition coming from one of the most renowned sociologists of today, Anthony Giddens, follows [Gid90]:

> Trust is related to absence in time and in space. There would be no need to trust anyone whose activities were continually visible and whose thought processes were transparent, or to trust any system whose workings were wholly known and understood. It has been said that trust is *a device for coping with*

> *the freedom of others*, but the prime condition of requirements for trust is not lack of power but lack of full information.

In 1994 the academic circles were captivated by the proceedings of artificial intelligence. Stephen Paul Marsh introduced the term trust into multi-agent systems [Mar94]. His understanding of trust was based on many humanistic definitions. Although his work is remarkably accurate and contains a complicated mathematical apparatus, the term *trust* in information technologies was never defined by him in an exact manner. The definition used from psychologist Deutsch follows:

> 1. The individual is confronted with an ambiguous path, a path that can lead to an event perceived to be beneficial (Va +) or to an event perceived to be harmful (Va -).
> 2. He perceives that the occurrence of Va + or Va - is contingent on the behaviour of another person.
> 3. He perceives the strength of Va - to be greater than the strength of Va +.
>
> If he chooses to take an ambiguous path with such properties, I shall say he makes a trusting choice; if he chooses not to take the path, he makes a distrustful choice.

2.2 Reputation

The term *reputation* is intuitively perceived as a general opinion on a social entity (person, group, organization), which is a result of assessing many factors. Reputation as such is never defined accurately in the humanities field. The best alternative is *social prestige*. It is explained by Šimíčková in [ŠČ04]:

> The important issue of each individual member of a group is his *prestige*, which is associated with his personal traits and his use in the group. It is the evaluation of the person by the group and it is called status (informally *reputation*, dignity, respectability [Kol01]). Status and position are connected and form a hieararchical system. We can encounter a system of statuses and corresponding social prestige in all people living in groups. The more traffic, the bigger the enterprise or institution, the more elaborate the system of hiearchy is.

Social sciences also recognize the term *position*. Position is important for trust contexts, their usage is further explained in 3.2. According to Řezáč [Řez98], the role which the individual occupies within the group is called a position. Positions define the rank to the group in dimensions of assignment, superordination and subordination. The position is determined by:

- the amount of individual's social appeal,
- the amount of prestige,
- the way of assertion within the group,

- the contribution to achieving goals of the group.

According to [Sch12] trust is a natural defensive mechanism of human society. Reputation helps us find out who to trust. When such an individual occurs in the society that misuses the trust of others and repeatedly acts in his own selfish interest, reputation is the best way of indication and warning others. Long-term unfavorable reputation leads to stigmatization and complete separation from the given social group.

There is another concept utilized in the intended implementation which needs explanation in the theoretical part. It is the fact that during the evolution of human society our central nervous system had no opportunity to adapt to new models of reputation. The so-called *Dunbar number* [Dun98] determines the number of people we are capable of maintaining social relationships with. According to research the number has been stable for a long time, contrary to the growth of today's society.

A new scalable formalism for reputation had to be formed in order to maintain the pace of development for this society. A parallel with the digital world may be observed here, since the scale rises drastically with the use of formalism and enables the use of trust and reputation in situations they were not intended for.

In the digital world trust has a meaning very similar to the physical one. Many internet communities are based on reputation, namely for instance Slashdot.org frequently used in trust research or the website eBay [Eba02], which directly uses reputation to evaluate individual users/sellers/customers. This particular example enables scaling beyond the originally intended borders, taking into consideration the greater risk of misuse.

Elizabeth Gray defines reputation in information technologies as follows [Gra06]:

> A reputation is a collection of recommendations, i.e., personal observations recommended by one or more third parties, about an entity's past behaviour which are accumulated in such a way as to characterise an entity's nature with regard to ability or reliability in potential future interactions in a given context. If the accumulated recommendations are evidence of behaviour for a given trust purpose, then the resultant reputation characterising an entity's trustworthiness can be used as input to a trust-based decision-making system.

This definition, however, is focused on trust, not reputation. To conclude this chapter, the difference between trust and reputation is presented. Doctor Audun Josang and his colleagues in [JIB07] expressed the difference nicely:

> The difference between trust and reputation can be illustrated by the following perfectly normal and plausible statements:
>
> 1. „I trust you because of your good reputation."
> 2. „I trust you despite your bad reputation."
>
> Assuming that the two sentences relate to identical transactions, statement (1) reflects that the relying party is aware of the trustee's reputation, and bases his trust on that. Statement (2) reflects that the relying party has some private knowledge about the trustee, e.g. through direct experience or intimate relationship, and that these factors overrule any reputation that a person might have. This observation reflects that trust ultimately is a personal and subjective phenomenon that is based on various factors or evidence, and that some

of those carry more weight than others. Personal experience typically carries more weight than second hand trust referrals or reputation, but in the absence of personal experience, trust often has to be based on referrals from others.

2.3 Social network

Prior to using social networks for creating a model for trust or reputation, it is necessary to define the term `social network` in the first place. This term originated in the widely quoted work of doctor Barnes describing his stay in a Norwegian village of Bremnes [Bar54]. In accordance with ancestral traditions and the isolation of this village from the other parts of the world Barnes was able to study certain class phenomena and categorize the inhabitants into certain groups. These relatively autonomous groups were then described with the term *social network*. After 50 years of its existence, the definition is as follows:

> A social network is a social structure made up of individuals (or organizations) called „nodes", which are tied (connected) by one or more specific types of interdependency, such as friendship, kinship, common interest, financial exchange, dislike, sexual relationships, or relationships of beliefs, knowledge or prestige. [Soc12]

For the sole purpose of comparison to modern perception of established terminology, this is the definition from the Oxford dictionaries [Oxf31]:

> *social network*, noun
>
> 1. network of social interactions and personal relationships,
> 2. dedicated website or other application which enables users to communicate with each other by posting information, comments, messages, images, etc.

As we can see in the second meaning, the term social network has for a certain amount of time been used to distinguish a network service used for communication and information exchange among users. This shifted meaning has a considerably larger dispersion.

2.4 Social network analysis methods

Having defined the social network, we also have to describe its analysis, which has been countinuously developed in the second half of the 20^{th} century. With the amplification of globalisation and mass production it started to make sense to analyse networks of customers and their needs. This particular time helped create a number of branches of science, i.e. new graph theory, sociograms, mathematical relation theory. A new kind of artificial intelligence emerged capable of predicting mutual relationships and needs of participants based on the collection and mining of data. The emerging field employed dozens of top-class mathematicians adept in statistics and data processing. The form of intelligence analyzing the nework is called *Numerati*. More information on this phenomenon can be found in the work [Bak09].

This thesis, however, needs key methods to analyze the social network itself. A brief overview of the most popular methods follows. They were summarized in the bachelor thesis [Mac11]. Each method contains an explanation why it is suitable or unsuitable for analysis in this work's particular context.

2.4.1 Complete analysis

Complete analysis requires collecting information on all relationships which occur in the network. It is equivalent to the Cartesian product in mathematics where for every two participants we need to express their mutual relationship. As expectable from the semantics of the name, this is the most complete and strongest type of analysis, which consumes the greatest amount of resources and time. For the particular purpose of this thesis, complete analysis would require contact with each individual member of the analyzed portion of the social network and their cooperation. We can only speculate as to how much of this type of analysis is being done in the background from the social network providers themselves.

2.4.2 Snowball method

Snowball method can be compared to the recursive query DNS system. It tracks down the relationships of an initial participant, finds new participants and recursively runs the same analysis for them. This calculation may be limited by either the amount of gained results or maximum elapsed time. The snowball method serves for analyzing connected communities within the social network, FBI allegedly uses this type to isolate terrorist cells [Bak09]. It is of course necessary to consider the disadvantages as well. This method cannot find isolated entities in the social network graph, as there is no way to guarantee all connections shall be found. An interesting problem to solve would be to find out whether the isolated communities tend to express more trust than the connected ones. This method seems to be suitable for the intended purposes, although it requires some basic understanding of the structural hierarchy of chosen network.

2.4.3 Community detection

Detecting communities, clusters and cohesive subgroups is one of the key tasks of social network analysis. This method analyzes groups of users based on the amount of traffic among them. Group detection can find valid criteria between nodes (reciprocity, availability or level). Symmetrical trust between nodes can according to some theories indicate the membership in a group.

2.4.4 Hierarchical clustering

Hierarchical clustering utilizes complex mathematical theories for creating clusters on several separate levels. It is not necessary to specify the estimated number of groups in the first place. This method, while widely used in social network marketing, is too complex for the application in the thesis and would interfere with the transparent use of suggested model.

Chapter 3

Trust model for Facebook

The aforementioned model constructed and validated in the bachelor thesis [Šve11] shall be redesigned in this chapter. Potential interested readers can find a detailed and original description of this model in the thesis.

Through the encouragement from doctor Samek the model was not put to rest after the bachelor thesis completion. A short summary in the English language was generated [ŠS13] and sent with an application for the UMAP conference 2013 [UMA31]. The paper was accepted for the trust workshop and conference participation was kindly funded by Brno University of Technology.

The week spent in Rome in June 2013 provided a number of prolific discussions about the model, culminating in the presentation in front of the leading experts on trust and reputation. The positive acceptance and encouragement from the community inspired further development of the concept and brought the redesigned version decribed in this chapter. For a short recapitulation, these were the intended properties:

1. keeping performance and resource needs at an acceptable level,
2. keeping the model scalable for future use,
3. no dependence on any sort of programming language,
4. no dependence on any sort of social network.

3.1 Enhanced model requirements

The requirements mentioned in 3 may be considered a general set of properties for a well-designed model in the academic circles. What is usually missing is the validation of fulfilling these requirements. Not just by enumerating reasons why these items had been correctly put into practice, but also perhaps formally assuring these facts. The next few paragraphs are an aftermath analysis of these requirements.

3.1.1 Performance and resources

As the analysis of a single Facebook profile took several minutes depending on the extensiveness of given content, several measures were taken in the design to improve performance and also consumed resources, the main part of which consists of needed bandwidth (the count would go as high as tens of megabytes). These are the main measures intended to reduce the resource demand:

- Reducing the timespan - repeating the analysis after several months with the same users reveals a certain inertia given by taking into account interactions that could not have affected trust so much.
- Omitting several contexts - as e.g. the context of common interests shows very little correlation in the final results and is the most demanding (taking into account time and bandwidth) it has been omitted completely.
- Utilizing server computation - some tasks performed in the Python script were also quite difficult to compute and took several seconds, utilizing server resources makes room for future utilization of more complex distribution of computational power.

3.1.2 Scalability

The suitability of the previous model to scaling has been tested in practice directly. As mentioned above, as some of the contexts needed adjusting, some contexts had to be omitted and some new were of course added, the workings of scalability may be visible in the newly formed model. At this point of explanation it is also appropriate to mention that the aggregation of individual trust contexts based on the explanation from Marsh [Mar94] remained the same in 3.5. The priority vector is, in fact, the part on which the scalability relies.

3.1.3 Programming languages

The original model was implemented in Python 3 and used Facebook Open Graph API. As it was created with the presumption the model would be implementable in any programming language, it is necessary to revalidate whether it is still possible to omit Python-specific functions and use a completely different language. This requirement had been validated, as the model was successfully implemented in PHP [PHP31]. As both PHP and Python are scripting languages, it would make sense to try to implement in an imperative language, for instance. On the other hand this would bring a lot of unnecessary effort caused by implementing web-friendly functions.

3.1.4 No dependence on any social network

This requirement is the only one which has not been taken into consideration. The present perfect tense is here to show that this fact does not have to be carved into stone, the effort to implement the model in another network would, however, be significant and probably not very productive.

3.2 Division into contexts

Marsh divides trust into so-called *contexts* in his work [Mar94]. These are based on areas in which we trust the given entity. To quote him directly: „Whilst I may trust my brother to drive me to the airport, I most certainly would not trust him to fly the plane!" This division of trust is absolutely crucial for determining trust among people, as it is a complicated subject simply not expressible by one single number in the initial phase. Using this principle may seem like too much of a generalisation, but when we put all the parts together, they enable us to work in a flexible and scalable manner (see 3.1.2).

The new model utilizes eight areas of trust, some of them in a way comparable to the original model, some of them completely new. A transparent overview is made into a table in 3.4. In these areas trust is observed between entities in the network. It was not possible to free the implementation from the specifics of Facebook completely, nevertheless the types of interactions still have counterparts which can be found in other social networks. To preserve the versatility of this solution, each context needs to be normalized. For simplicity a closed interval has been chosen from 0 to 1. A list of contexts and their short description follows. Many insights originate in the work [Sam11].

3.3 Context overview

3.3.1 Durability of friendship

- **Meaning:** This context is measured by the criteria of how long the entities have maintained a friendship on the social network. The longer I have known someone, the longer time I had to build trust with him. While intuitively the first variable to measure, the information value is quite low. A number of things could have happened in the timespan, both beneficial and malicious for trust. Therefore this context is given little priority.

- **Implementation:** The implementation is actually a little tricky in this case, as Facebook Open Graph API does not provide access to this particular piece of information. A workaround was necessary when a current list of actions is examined. This list also includes messages on the newly formed friendships and the date of such an event. If such an event does not occur, we assume we have been friends with this person for the whole analyzed timespan.

- **Normalization:** The resulting value is divided by the timespan length of the overall analysis, resulting in a number between 0 and 1.

3.3.2 Amount of contact

- **Meaning:** Amount of contact is one of the key parts of this model. *Contact* is a one-sided compact stream of information, speaking in the language of Facebook it is a wallpost, a comment or like. It is important to distinguish contact and a message - messages are processed in a separate context. There are users in the social network who are more active than others. These have to be filtered by setting a limit above which trust in this context simply cannot reach.

- **Implementation:** The algorithm takes into account comments and likes on analyzed user's wallposts. Each of these interactions is assigned a creator ID from Facebook API and this ID is used to access a list of user's friends. Wallposts also comprise newly published photograps, which gives us access to almost almost all activity a user participates in.

- **Normalization:** Normalization can be described by two simple equations

$$A = \frac{1}{n} \cdot \sum_{x=1}^{n} I_x \tag{3.1}$$

$$T_N(x) = \frac{I_x}{A + \frac{1}{n} \cdot \sum_{x=1}^{n} |A - I_x|} \tag{3.2}$$

where I_x stands for the number of interactions with a person x, A is an average number of interactions and the factor in the second equation is a sum of average counts and absolute deviation of all gathered values. This calculation was determined empirically and offers the most accurate results.

3.3.3 Regularity of contact

- **Meaning:** Regularity of contact differentiates users who communicate sporadically and users we communicate with on daily basis. Situations may occur where we communicate with one person excessively (preparing a celebration, concert, etc.), while not being particularly close to them. It is a key understanding in this context that it is only a *part* of the model and when used out of context it may present peculiar results. Other contexts are needed to amend this behaviour.

- **Implementation:** The context utilizes the metrics used in [ZTL+11]. The metric converts intervals between interactions into seconds and then multiplies all the values. The given product is maximized for equally distributed interactions in time. The equation goes as follows:

$$x_v^{\Delta T}(A, B) = \prod_{i=1}^{n-1} |t_{i+1} - t_i| \tag{3.3}$$

where n is a natural number and stands for the number of intervals taken into consideration and t is the time of interaction. The important restriction is the interval being greater than 1. TrustNet [ZTL+11] calculates this value in milliseconds, which on the other hand would result in numbers too big in this case. It is important to note that PHP offers much greater possibilities than previously used Python, mainly because of the possible simple utilization of server computational power. Big numbers produced by this formula are no problem for PHP.

- **Normalization:** The maximum value achievable for the highest count of interaction that occured is computed. The number is then used as the 100% value and other values are calculated proportionally. One additional aspect is the non-linear dependency of results. These are therefore reduced by the square root of the number of used interactions.

3.3.4 Real-life experience

- **Meaning:** Real-life experience has to be documented in a way to be taken into account in this model. Common photographs are the way to go in this particular case. It may be considered an overlap of this model into the physical world, because the photograph indicates a personal encounter of participants. This context is therefore given greater importance. There is a certain danger of distortion by photographs such

as Christmas wishes or jokes where multiple users are tagged. The solution is simple, setting a limit and excluding the photographs exceeding the headcount.

- **Implementation:** In principle this context is calculated from metadata in photographs. There are two sources of photographs the model can rely on. One of them is the stream of photographs the user has put onto his wall and which do not appear in his albums. The other one is the list of album photos. Both these types provide information for answering the simple question "Who was the user tagged with?".
- **Normalization:** The list of all friends is then sifted and the maximum reached value is stored. This value means trust of 1 in this context and serves for equal distribution for the remaining users.

3.3.5 Allegiance to communities

- **Meaning:** It is crucial for this context to describe the term *community*. In the beginning of Facebook, *groups* were a tool for spreading jokes, advice, advertisement. The shift of meaning still has not found its roots in users today. Groups or communities on Facebook today are a means of sharing information with a group of people. A few examples are people who are working on a project or commuting to the same city. These groups are frequently invite-only. There is an inverse relationship between the size of groups and their importance to trust. A commuting group for 250 users is not as significant as a group of 5 people working on a common project.
- **Implementation:** The groups of the analyzed user are gathered and then his friendlist is analyzed. For each friend his or her groups are matched to those of the analyzed person. Then an inverse value is calculated for importance based on the number of participants in this group. Without the use of batched requests, this context would be very time-consuming to compute.
- **Normalization:** Normalization is the same as in the context of Real-life experience 3.3.4. There is one additional aspect, groups with a very high number of members can be omitted as they do not contribute significantly to the resulting value.

3.3.6 Contact preferences

- **Meaning:** The concept of the Dunbar number was explained in the chapter 2.2. According to Bruce Schneier in [Sch12], the people who have common interests and pesonal traits tend to trust each other more. The Dunbar number is a value encoded into our neocortex and should vary between 100 and 230 according to current research [Dun98]. The deviation from this number is a trait which should bind similar types of people together and indicate more trust between them.
- **Implementation:** The implementation of this concept is still not possible due to Facebook's restrictions. Apparently the amount of friends one user has is confidential information and for security reasons can only be found for the current user, not for his friends. Therefore a simple alternative was chosen - the amount of *mutual* friends two users have. This context is the only one which utilizes the FQL API [FQL11].
- **Normalization:** The application finds the greatest number of common friends and takes it as the maximum value. This value means trust of 1 in this context. The rest of the values is equally distributed.

3.3.7 Amount of one-to-one contact

- **Meaning:** Originally this context was not used in the first model. Questionnaires for the users participating in the research showed, however, that this is considered the most revealing context of all [Šve11]. Personal data protection shall be explained on the website introducing the application. The idea behind analyzing the age of messages is quite clear - the more often we communicate with someone, the more we trust them. See the explanation in 3.3.2. It is crucial to distinguish interaction and messages, as the model should remain scalable and in case of personal data protection issues this context can be taken away.

- **Implementation:** Another simplification is necessary in order to remain in relatively manageable computational time. Real number of messages would take tens of minutes to compute. The last message interaction with given user is taken into consideration instead.

- **Normalization:** There is an inverse relationship between the amount of time since the last message and the amount of trust. Immediate contact is considered as value 1, no contact at all stands for value 0. It is therefore not possible to achieve number 1, as the analysis lasts nonzero amount of time.

3.3.8 Events participation

- **Meaning:** Again, this context is very similar to the one described in 3.3.4 and its explanation also holds. The key difference here is attendance of a formal event. As the offer of event-planning tools widens on Facebook (i.e. the provision of maps indicating the place directly on the Facebook website etc.), events are being planned in a more extensive manner. Common attendance means a probable real-life interaction, although it is not as sure as a photograph and therefore should be given less importance in the priority vector.

- **Implementation:** All Facebook events attended by the analyzed user are taken into account and crossmatched with the events attended by his or her friends in the given period. A certain limit has to be set in order to eliminate events with too many participants.

- **Normalization:** The process of normalization is identical to the one described in 3.3.4.

3.4 Comparison to the original model

One very important insight has to be incorporated into this work. As pointed out by the audience in UMAP 2013, there is a semantic misunderstanding in the original concept of *contexts* in the bachelor thesis [Šve11]. While the trust is being computed correctly, the names used in the thesis are not contexts per say. The terminology is used in the wrong way.

As an example, we take the Number of interactions. Number of interactions is not a context, it is a *criteria* to assess a context of Amount of contact 3.3.2. Using the criteria we estimate how much we can trust a person in the particular context. This conversion shall be shown in the following table together with overall changes in the included contexts:

Original name	New name	Reason for change
Interaction timespan	Durability of friendship	Criteria vs. context
Number of interactions	Amount of contact	Criteria vs. context
Interaction frequency	Regularity of contact	Criteria vs. context
Common photo tags	Real-life experience	Criteria vs. context
Common group membership	Allegiance to communities	Criteria vs. context
Number of friends	Contact preferences	New context
Number of messages	Amount of one-to-one contact	New context
Event participation	Events participation	New context
Common interests	-	Performance issues
Character count	-	Not enough info value

Table 3.1: Comparison of used contexts.

3.5 Context aggregation

The described contexts exhibit a degree of correspondence to real-world emotions, more of which is described in chapter 3.2. A more difficult task remains to aggregate these contexts into one single value expressing trust. Marsh in his ground-breaking work [Mar94] aggregates contexts by the method of simple multiplication. Multiplication does not, however, take into account the importance of individual contexts. As some emotions contribute to the result in a more specific way, it was apparent from the initial design that contexts would have to be weighed in a certain way.

The solution to the stated problem lies in the *priority vector*. In a mathematical sense it is an ordered set of coefficients which takes into account the importance of individual contexts. To express each of them, an abbreviation was chosen for the equations. The abbreviations are recorded in table 3.2.

Full name	Abbreviation	Priority
Durability of friendship	DoF	1
Amount of contact	AoC	3
Regularity of contact	RoC	4
Real-life experience	RlE	1
Allegiance to communities	AtC	2
Contact preferences	CoP	1
Amount of one-to-one contact	AoO	4
Events participation	EvP	2

Table 3.2: Abbreviations chosen for contexts.

These values then represent trust in separate contexts:

$$P = (T_{DoF}, T_{AoC}, T_{RoC}, T_{RlE}, T_{AtC}, T_{CoP}, T_{AoO}, T_{EvP}) \tag{3.4}$$

The coefficients in table 3.2 are used to aggregate the mentioned trust contexts into one single value T determined by the following equation (triplets of letters represent priority as in the table):

$$T = \frac{\sum_{x\epsilon\{DoF,AoC,RoC,RlE,AtC,CoP,AoO,EvP\}} T_x}{DoF + AoC + RoC + RlE + AtC + CoP + AoO + EvP} \tag{3.5}$$

It was stated in the original thesis [Šve11] that the priority vector is capable of withstanding changes to individual contexts' priority. This proved to be true in the sense that it has been significantly altered and still enables the same functionality for not only altered priorities, but also new and altered contexts. The reasons for setting priorities are described in section 3.3.

Chapter 4

Personality psychology

As stated in chapter 1, one of this thesis' main contributions to the trust and reputation field is its utilization of the humanities field in connection with information technologies. The core of this thesis lies in verifying the results of the proposed model with psychological data from real social network users. As the intersection of fields may be an issue here, precautions have been taken and a few facts must be stated in advance.

First of all, it is not humanly possible (and neither desirable in this case) to introduce the whole field of psychology to the common reader. This text is constructed in a top-down approach and introduces new terms in the order in which they are needed to get to the corresponding level and to explain the theory behind the questionnaire 5.2. The facts and definitions are brief because full explanation would require more space than a master's thesis offers. Initiative consumers of this text may find further information on this field in used literature, particularly recommendable is the publication from doctor Kohoutek [Koh02].

Secondly, most sources used in this chapter are originally composed in the Czech language. The reason for the choice of such sources lies in the simple fact that a Czech police psychologist was consulted in the creation process and the key definitions used to theoretically approach the subject are more or less very similar in both languages.

4.1 Initial model requirements

At the beginning of this thesis stood a goal of deeper understanding for factors that influence the accuracy of the constructed multi-context trust model for social networks. It is natural that such a task requires analysis of the human beings as active users of the social network (and correspondingly the trust model). There were also requirements for the theoretical basis used to compare individual users:

1. User comparison should not depend on the momentary mood or state of mind of the subject, i.e. the observed traits should be stable.

2. The chosen method of comparison should not be too complicated, otherwise it would not be possible to isolate any correlation to trust.

3. The method should be easily implementable in computer-readable form and ideally should also utilize only mathematically simple calculations.

4.2 Personality psychology

The field in which the answers to given requirements shall be found is the personality psychology. Unless stated otherwise, this section draws facts mainly from doctor Kohoutek's publication *Basic of Applied Psychology* [Koh02].

Personality psychology is one of the most important branches of psychology and its primary subject of study is personality. Its results are used in meeting and negotiating with people, personal activities (choosing employees), solving interpersonal conflicts, upbringing and imperative work, solving personal issues of employees, etc. *Personality* is most easily and intuitively defined as:

- what we *want* (instincts, needs, interests, values),
- what we *can* (abilities, predispositions, talents),
- what we *are* (temperament, character).

Etymologically the term personality comes from the latin word *persona*, which originally meant a theatrical mask used by actors in ancient plays to change and indicate their roles in the play. As far as an accurate definition is concerned, doctor Kohoutek defines:

> Personality can be defined as dynamic organisation, summary, unity of inherited, congenital, relatively stable pecularities created under the pressure of society and education (and self-education) in interaction with natural, social, economical and cultural environment, bodily, bio-psychological and spiritual processes, attitudes, relationships, dimensions and characteristics, which determine the activity of man and condition not only his behaviour, but also experiencing and adapting to the environment.

The whole field of personality psychology as a bio-psycho-social system can be divided into three basic areas:

- **personality dynamics** which explores and categorizes all forces activating experiences and behaviour,
- **personality development**, onthogenesis, explores characteristic behaviour and experiences in individual life phases vertically, and
- **personality structure** which identifies components that compose personality and their inner architecture.

The structure of one's personality is the factor that is most interesting for the thesis, as it offers stable results, which is according to requirement number 1 in 4.1.

4.3 Personality structure

Personality structure shall be defined here using the publication from doctor Pauknerová, *Psychology for Economists and Managers* [Pau07]. As this work is more oriented into the practical field of psychology, it allows an insight into the usability of individual dimensions of personality and therefore lets us choose the dimension most suitable for the intended

utilization. *Personality structure* itself is defined as everything relatively stable in one's personality, includes its characteristics and their relations. Related characteristics are categorized into three groups which we call dimensions (each of them shall be described in its individual subsection):

- performance characteristics,
- motivation characteristics,
- profiling characteristics.

4.3.1 Performance characteristics

Performance characteristics are in direct connection with a person's performance and his/her practical application in life and work in particular. The amount of one's success performing any sort of activity is influenced by a group of factors, namely biological, social and self-forming factors. Inner factors are described as endowments. They represent specific anatomical and physiological characteristics of an organism. If several endowments meet in a similar area, the individual posesses a talent. Further levels include abilities, knowledge and skills. This dimension, however, is more connected to the description of an individual concerning work. It is not suitable for correlation with trust.

4.3.2 Motivation characteristics

The level of performance is also measured within the borders of motivation. Low motivation also indicates low performance levels, while the case of exceeding motivation is also very similar. Motivation in general describes why a person is active and why he/she behaves in a particular way. It is usual for motivated behaviour to aim at a certain goal, invest a certain amount of energy corresponding to the desired goal and the duration of motivation, which is usually determined by the reaching of the goal.

There are two basic sources of motivation - inner and outer, which originates in the interactive nature of motivation. The human behaviour is determined by his/her inner state. Inner sources of motivation are described as motives (i.e. thirst), outer sources as stimula (i.e. water). Abraham Maslow [Mas43] takes the principle of stimula and transforms it into so-called needs. The needs are categorized in the famous Pyramid of needs portrayed in figure 4.1.

According to these facts, motivation could indeed exhibit some correlation to trust in social networks, i.e. for people who are highly motivated to get social approval by peers and therefore are more active. There are two problems regarding this approach. First, it is virtually impossible to take into account all sources of motivation one can possibly have. Second, the numerous groups of motivation are very difficult to work with and thus do not comply to the requirement 2 in 4.1.

4.3.3 Profiling characteristics

As the previous two dimensions indicate, this is the chosen dimension for forming a model and assessing the accuracy of trust computed for individual users. Profiling characteristics are the most stable of all dimensions and are perhaps most widely known in public, which provides a certain degree of motivation for potential users who participate in the real-life research on social network. There are three characteristics present in this dimension and

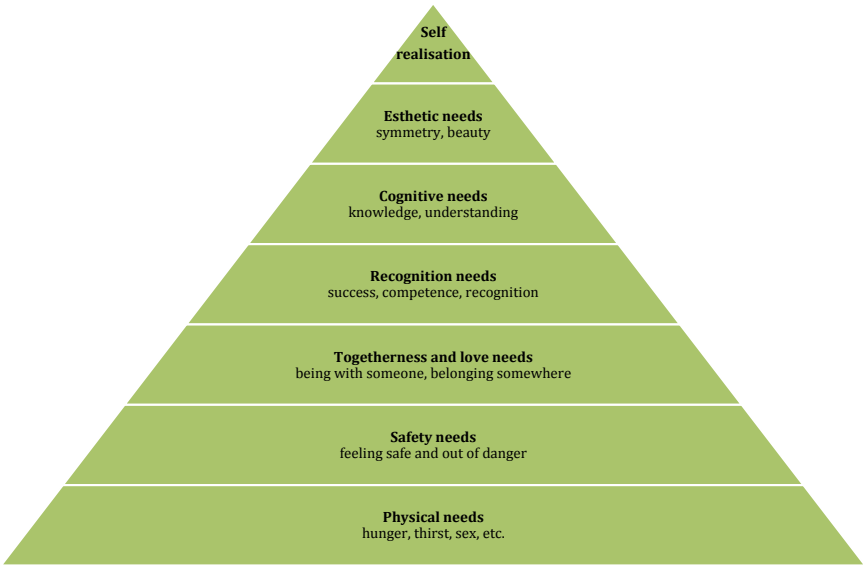


Figure 4.1: The well-known Maslow pyramid [Mas43].

each of them would be very interesting for finding correlation between two models. This, however, would sadly exceed the given space for a master's thesis.

Temperament

Temperament expresses a formal form of our experiencing and behaviour, it is currently assigned to individual needs of stimulation, outer triggers. Temperament is a characteristic of personality which is most biologically conditioned, meaning we are born with it and can only mold it, but not completely change it. Exhibits of temperament can be traced to first months of life and are most easily identified in emotionally tense situations.

Temperament is generally divided into four types - choleric, melancholic, sanguine and phlegmatic. Even though this division has been widely used since the antics, the perception of the source of characteristics keeps changing dynamically. Currently the most accepted model comes from Eysenck and combines two basic dimensions:

- **stability/lability** of the nervous system,
- **extraversion/introversion** to other people.

These two dimensions form the well-known shape used in many online tests for temperament and portrayed in figure 4.2.

As far as usability is concerned, it would be very interesting to observe how nervous system affects people's behaviour on social networks. The main issue here lies within the simple fact that temperament can easily be influenced by learnt behaviour. It would take a complicated test to analyze users' temperament and they probably would not donate so much time to the research.

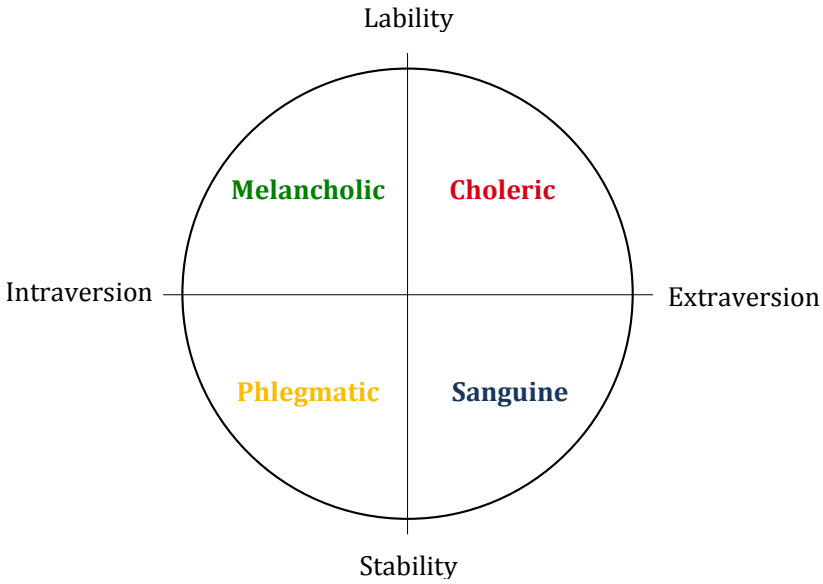


Figure 4.2: Temperament cross [Shi98].

Attitude

Attitudes are another significant profiling characteristic of personality. They are relatively stable judgements, opinions, approaches, emotional relationships and tendencies to act similarly in similar situations. They stabilize each person's individual experiencing and behaviour. Some attitudes are expressive, are intensively experienced and are clearly reflected in human behaviour. Others are not so significant and affect „smaller" opinions. The more important the matter is for a person, the more problematic forming an attitude may be. They are formed and developed above all:

- after a significant number of the same or similar experiences,
- as a result of a very intensive one-time experience, a strong emotional event, i.e. endangering of life,
- by transfer from personally close people or people with very similar opinions.

As can be deduced from the above definition, attitudes are *learnt over time*. They reflect values held by people and are thus connected to motivation. They form an inner interconnected system and relationships can be found and traced. This is also the reason why they cannot be easily changed.

It is this last fact that makes our work more difficult, should we choose attitudes as comparing characteristic for the trust model. It has been stated by several people at the UMAP 2013 conference that attitudes have a significant impact on trust mainly in the initial phases of its development. Attitude is a very prolific topic to be discussed in connection with trust, the main issue here was the complexity and interconnection among all the attitudes of a single human being. The requirement number 3 in 4.1 is therefore not satisfied.

Character

Character is used to express relatively stable characteristics of personality, which are formed and also occur in one's relationship to parts of reality, including the particular human being. It is connected to morals and ethics, eventually maintaining generally accepted moral principles. Character properties are both stable and variable for particular parts of personality and are formed in a major way in the process of socialization, in childhood and youth.

Character offers the most desirable properties for the given trust model. It represents both stable and variable parts of personality, but is not susceptible to momentary mood swings. It is not a complicated characteristic. And, most of all, the models researched to be in correlation with character are very easy to work with and are not difficult to compute, see 5.2. Thus the requirements stated in 4.1 are fulfilled and we can proceed further by describing exactly what character is in its own section 4.5.

4.4 Hierarchy overview

As the presented field is quite vast and very easy to lose overview in, the following hierarchical diagram shows the elapsed path from personality to character before describing character itself in its individual section 4.5.

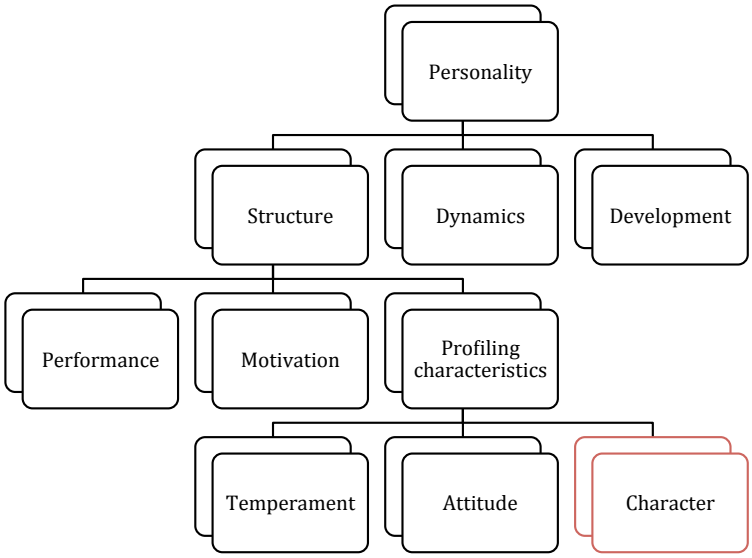


Figure 4.3: Personality hierarchy overview [Koh02].

4.5 Character

A majority of this section, unless stated otherwise, draws information from the source by doctor Kouhoutek [Koh02]. His definition is far more precise than the one used in people

management by [Pau07]. In the appropriate places the definitions are commented with respect to the utility in the intended model of trust.

Many psychologists do not consider character as being part of psychology, because they find it too normative and demolishing the empirical nature of this science. In this context character is not considered a synonym to personality, it is considered only its part, the moral core with the essence in *conscience*. The properties of character manifest on the outside by relatively stable behaviour to both the outside world (social in particular) and to one's self. Outer manifests of personality characteristics can be expressed in the terms of *streaks* (traits of personality), such as truthfulness, diligence, honesty, sociability. The equivalent to the term *character* is *nature*.

Character was first mentioned in Theofrast in the third century BC [The10]. The term originally meant a seal and coin minting (*charassein* is Greek for *carve* or *print*), in a metaphorical sense it was meant as a distinctive sign. In our sense we understand character as the behavioral core of personality and a set of individual traits, which differentiate the person from others. These traits influence the behaviour and relationship of said person to himself, to society, work, they grasp the quirks of his free will. If we know these traits, we can (with certain precision, of course) predict the person's behaviour [Rub64]. The value of person is determined by the inclination to relationships, which relationships he maintains or seeks. The connection to trust is palpable here. Character (or the person's *orientedness*) is gained to a certain degree, it is influenced by learning processes, education, family, school, society, later even on self-education. Character manifests itself in and also molds the way of life. The most intuitive definition by doctor Kohoutek follows:

> Character can be defined as tendency to behave and react in a certain way in a certain situation.

This is exactly the question we are asking ourselves when computing trust. How are the actions which were taken into account in the model actually influenced by a person's character, the tendency to react to social interactions in a certain way?

4.5.1 Character in relation to other people

This is obviously the context which is most interesting for us. The quality of relationships to other people is determined by personal morality (altruism or egoism, the sense of good, evil and justice). Altruistic, non-selfish and friendly people live by moral rules, egoists do not care about the opinions of others, should they clash with their individual interests. It is also vital to assess the so-called *social intelligence* or social sensitivity, social tact, e.g. *social hostility*.

We assess the attitude to men, women, co-workers, classmates etc. We also observe character traits typical for both individuals and their corresponding group. The description of character traits in relation to other people should encompass the *dominant* and *submissive* inclination to men, women, elderly and peers, the degree of *expansivity*, *sociability* or loneliness, popularity and conflictivity. We are also interested in the degree of *hypocrisy*, which is typical for adults. Children tend to be spontaneous and direct. It is also remarkable that unsociable people consider unsociability more indulgent than the sociable people.

4.5.2 Character in relation to self

Although the relation of character to one's self does not provide as much studying material as the relation to other people, the self-perceived personal value is still a contributing matter

in trust. When a person is unaware of his or her personal value, his or her worth, he/she is often able to publicly ignore injustice or wrongdoing on self or others. Low self-esteem often leads to reduced motivation to learning and reaching personal goals. Positive self-assessment encourages activity, stimulates utilizing opportunities and also risking defeat. This includes participation in social activities both in real world and in a social network.

There are several types of self-assessment emerging in adulthood that are developed since early childhood. Many people cannot be categorized into exactly one of these types accurately. A majority of people is susceptible to changes in self-assessment, which leads to unstable self-assessment. It is after all affected by the appreciation from the environment, popularity, sympathies, admiration, recognition etc. These people possess a so-called *reflexive* (mirroring) self. This type of self is generally considered more common in women than men. It is also more common in lower age groups.

The reflexive self is formed based on the perception and understanding from other people, how they appreciate our personality, our expression and behaviour, temperament, character, abilities. It is the reaction to assessment by the surroundings. In connection to this evaluation one can feel pride, satisfaction or detunes, shame, feeling of inferiority. Reflexive self is the genetic factor of *conformism*. People generally consider assessment from the surroundings to be important for their life's success and professional career. In other words, the reflexive self is the key motor to activities in social networks.

Chapter 5

Big Five and its relevance to trust

The previous chapter 4 explained why the usage of character is the best metric which can be used in this particular case. This chapter explains how the most suitable model for representing character was chosen, the method of acquiring user data on this matter using several types of questionnaires, and last but not least the expected correlation between the chosen model of character and the created model for trust. This expectation was foreshadowed in the preceding term project and shall be analyzed in detail in this continuation.

5.1 Character models taken into consideration

There are various ways how to express character and translate its complex nature into a set of mathematical values which can then be used to compare different personalities in different people. It is this particular field where informatics and humanities come close together. Sometimes information technologies offer the best way of processing data for humanities, sometimes, like in this concrete case, humanities provide inspiration for information technologies and these two work together. The following models were taken into account when choosing the most suitable model for character used in this thesis. The descriptions are very brief and the most space is dedicated to the chosen model in the next section. The chosen model called Big Five is given its own section 5.2. Information on the Big Five model and also for the general overview utilizes [Svo05] as its source.

5.1.1 16 Personality Factor Questionnaire

The 16 PF Questionnaire had been in research and in continual improvement by its creator, R. B. Cattell, for a number of decades [Gre11]. While today quite deprecated and overcome by newer forms of research, this questionnaire represents the foundations of modern psychological methods. Factor analysis was used in order to measure the source traits of human personality. The 16 factors in table 5.1 are observed. This model, while very accurate, lacks the simplicity for finding any real connections between the models' parts.

5.1.2 Freiburger Persönlichkeitsinventar (FPI)

The FPI is a multi-dimensional questionnaire which was published in 1970 by J. Fahrenberg, H. Selg and R. Hampel [Gre11]. It was used mainly in German speaking countries and has more connection to clinical psychology than any questionnaire described in this chapter.

Warmth	Reasoning	Emotional Stability	Dominance
Liveliness	Rule-Consciousness	Social Boldness	Sensitivity
Vigilance	Abstractedness	Privateness	Apprehension
Openness to Change	Self-Reliance	Perfectionism	Tension

Table 5.1: Factors in the 16 PF Questionnaire.

The diagnosed dimensions are described in table 5.2. FPI is already nearing our set of requirements, the only issue is its aim at personality disorders, not ordinary traits.

Number	**German factor**	**English translation**
1	Lebenszufriedenheit	Satisfaction with oneself
2	Sociale Orientierung	Social orientation
3	Leistungsorientierung	Need for achievement
4	Gehemmtheit	Shyness
5	Erregbarkeit	Irritability
6	Aggresivität	Aggressiveness
7	Beanspruchung	Demandedness
8	Körperliche Beschwerden	Physical troubles
9	Gesundheitssorgen	Health sorrows
10	Offenheit	Openness
11	Extraversion	Extraversion
12	Emotionalität	Emotionality

Table 5.2: Factors in the FPI Questionnaire.

5.1.3 Personality questionnaire by B. Miglierini

This questionnaire represents the original approach to personality measuring in Slovak literature [Gre11]. It diagnoses 20 personality traits and clusters them into 8 factors. This clustering technique is what will be interesting for the Big Five model. It takes about 30 - 45 minutes to fill out the questionnaire and it is therefore not suitable for the intended purposes. Factors taken into consideration are described in table 5.3.

Number	**Factor**
1	Sociability and activity
2	Independence and decisiveness
3	Emotional stability
4	Family relations
5	Physical and mental welfare
6	People relations
7	Interest orientedness
8	Conscientiousness and responsibility

Table 5.3: Factors in the questionnaire by B. Miglierini.

5.2 Big Five

5.2.1 History and origin

Roughly twenty years ago a new personality description system emerged which has gained popularity throughout the world and is considered standard in the Czech Republic as well [Říč07]. Big Five is a „descendant" of Cattell's 16 PF (see 5.1.1). Its authors, Ernest Tupes and Raymond Christal, used cluster analysis together with the 16 PF data and reached three clusters of Cattell's primary factors. These three clusters were *Neuroticism*, *Extraversion* and *Openness to experience*. They became the basis for the creation of a new questionnaire.

The remaining two factors, *Agreeableness* and *Conscientiousness*, were amended based on lexical studies. Variables corresponding to these factors were created and these variables were factored together with the original triad. The five factors were independently discovered by several other researchers whose research was based on scales factoring.

5.2.2 Big Five principles

Adjectives which are used for describing personality were studied by researchers based on general lexical hypothesis [Svo05]. Several independent researchers concluded the same five orthogonal factors:

- **O - Openness to experience** (or Intellect) - evaluates active search for new experiences, tolerance to the unknown and its discovery.
- **C - Conscientiousness** - individual level of organization, motivation and aim for achievement. It differentiates people who are reliable or indifferent and negligent.
- **E - Extraversion** - explores the quality and quantity of interpersonal interactions, the level of activation, the need for stimulation.
- **A - Agreeableness** - measures tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others. It is also a measure of one's trusting and helpful nature.
- **N - Neuroticism** (or Emotional stability) - measures the level of adaptation of emotional instability, neuroticism. It can help tell apart individuals prone to mental exhaustion and unrealistic ideals from individuals who are balanced and resistant to mental exhaustion.

The five factors form an acronym: *OCEAN*, which is a term generally used in Big Five literature. Personalities analyzed with the Big Five model are ranked with adjectives corresponding to the low or high level of the particular factor in the observed personality, see table 5.4. The values are generally normalized in the closed interval from 0 to 1.

5.2.3 Big Five questionnaire

In contrast with the commonly used Cartesian four-dimensional model, the five-dimensional model of personality traits needs a more empirical approach. Data is usually collected in a questionnaire presented to the subject. There are many types of questionnaires we can encounter in this area, i.e. NEO-PI (NEO Personality Inventory) from Costa and McCrae

	High level	Low level
O	curious, original, imaginative, creative, versatile interests, progressive, intelligent	conventional, narrow-minded, non-analytic, non-artistic, conservative, non-intelligent
C	hardworking, disciplined, accurate, punctilious, tidy, self-challenging, persistent	aimless, unreliable, lazy, slobby, indifferent, without will, mammonish
E	active, talkative, optimistic, funny, people-oriented	closed, serious, silent, task-oriented, quiet
A	good-hearted, kind, trustworthy, helpful, sincere	cynical, crude, distrustful, non-cooperating, revengeful, heart-less
N	tense, restless, unsure, nervous, labile, hypochondric	calm, relaxed, balanced, stable, self-conscious, satisfied, unstrung

Table 5.4: Adjectives associated with OCEAN levels.

[MC85], FFPI (Five Factor Personality Inventory) from the Netherlands led by Hendriks [Hen09] or the BFQ (Big-Five Questionnaire) by Caprara [CBB11].

All the previously mentioned questionnaires have one ill. They comprise over a hundred adjectives which need to be self-assessed by the correspondent in order to evaluate the Big Five model. For this particular reason a different method was chosen. This method was developed by professor George Boeree at Shippensburg University and utilizes only 40 representative adjectives [Boe09]. Although the accuracy could suffer in a minor manner here, the agile approach allows us to present the questionnaire to a mass of people who do not have enough time to fill in a large amount of data. The questionnaire was implemented in PHP as part of the thesis and forms one part of the resulting application. Its description can be found in section 5.4.

5.3 Correlation with trust

Correlation with trust is the most important section of this chapter and it is also the reason why Big Five was taken into consideration. Based on the so far gathered facts, the personality of the observed person should affect the way trust is formed in them. It is yet unclear what the precise relationship is. We can try to make a qualified guess what the relationship will look like (and the qualified guess follows), the comparison shall be processed based on actual questionnaire data in section 7.6.

The contexts presented in chapter 3.2 are compared to the OCEAN factors and expected correlation is explained in the process. The section is concluded with an overview in the form of the table 5.5 which summarizes the expected correlation and also offers a general view of the whole process.

5.3.1 Context relevance

1. **Durability of friendship** - The first context is a bit special. It does not utilize openness to experience and is not very dependent on extraversion either. The two main factors to be taken into consideration here should be agreeableness and neuroticism, because it takes a *stable* and *cooperating* person to maintain friendships.

2. **Amount of contact** - Amount of contact has very much to do with openness to experience and most of all extraversion. As the meant interactions are seen by the whole community of friends, only *talkative*, *people-oriented* and *curious* people post their thoughts in a larger scale.

3. **Regularity of contact** - This context utilizes the whole spectrum of the Big Five model. Factors most important to this context are again openness to experience and extraversion. The expected inclination to experience is based on the term *versatile interests* used in table 5.4.

4. **Real-life experience** - Without much doubt, real-life experience represented by common photos requires a certain level of extraversion to be present, as well as openness to experience. The most important adjectives here are *active*, *talkative* and *people-oriented*.

5. **Allegiance to communities** - Membership in communities is not a usual context. It requires a different factor to be utilized, namely agreeableness, next to openness to experience. Allegiance to communities is traceable in people who are *trustworthy* and have *versatile interests*.

6. **Contact preferences** - Again, this context is dependent on openness to experience and extraversion, agreeableness is also a factor. People with a high number of friends tend to be *curious*, *non-conservative*, *talkative* and *people-oriented*, as well as *trustworthy*.

7. **Amount of one-to-one contact** - Interestingly enough, this context does not seem to have any relationship to the OCEAN factors which could be predictable. As the contact is private, there is very little to say about how it utilizes the Big Five model. Results will have to be compared in the collected data.

8. **Events participation** - Very similar to real-life experience, this context utilizes the same factors.

5.3.2 Table overview

The expected relevance has been rated on scale from 0 to 3 with the following meaning:

- 0 - none,
- 1 - existent,
- 2 - high,
- 3 - crucial.

5.4 Questionnaire user experience

The final section of this chapter deals with the problem of the questionnaire implementation itself. The questionnaire's pages implemented in HTML generated by PHP will be briefly described and user experience also narrated. As the implemented site serves for the interface

Context	Relevant factors				
	O	**C**	**E**	**A**	**N**
Durability of friendship	0	1	1	2	3
Real-life experience	2	0	3	1	0
Contact preferences	2	0	2	1	0
Allegiance to communities	2	1	1	3	0
One-to-one contact	0	0	1	1	1
Events participation	2	0	3	1	0
Amount of contact	2	0	3	1	0
Regularity of contact	3	1	2	1	1

Table 5.5: Overview of trust and expected OCEAN correlation.

with users, this is perhaps the part of the thesis which wanders most from the information technologies field. On the other hand it seems vital to describe the purpose of each of the modules implemented for flawless user experience. For future complementary reference, the technical implementation is described in subsection 6.2.9.

5.4.1 Introduction

In order to get a higher number of participants, the website was implemented in two languages. The introduction page serves only for choosing the language. It has to be simple, yet eye-catching. Two simple flags and the application names in their respective languages are divided by a figure trying to decide which flag to take. Screenshot 5.1 shows the final look with the applied CSS style. The final design of the page is also nicely aligned in the sharing thumbnail in social networks.

Figure 5.1: The welcoming screen of the questionnaire.

5.4.2 Big Five questionnaire

The solution presented by doctor Boeree in [Boe09] had to be enhanced in one more way. It is again a sort of generalization for the common population. Doctor Boeree uses fitting adjectives for users' self-evaluation, which describe their corresponding factors in the OCEAN model. This notion is somehow abstract and not all people are capable of fully grasping the meaning of some of the adjectives. They were therefore transformed into simple sentences describing the same characteristics. As an example, we can look at the word "creative" replaced with the sentence "I am full of ideas". The list of used sentences for each individual OCEAN factor follows. An example of a questionnaire screen can be seen in figure 5.2.

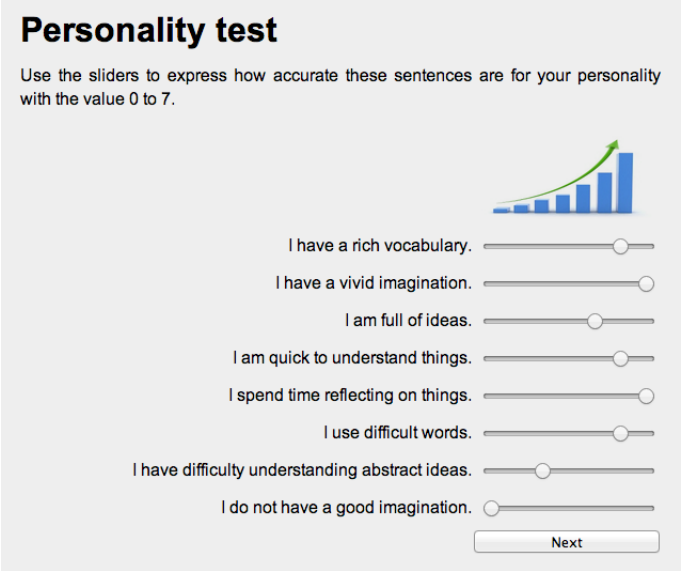


Figure 5.2: One of the five screens of the questionnaire.

- Openness to experience
  - I have a rich vocabulary.
  - I have a vivid imagination.
  - I am full of ideas.
  - I am quick to understand things.
  - I spend time reflecting on things.
  - I use difficult words.
  - I have difficulty understanding abstract ideas.
  - I do not have a good imagination.
- Conscientiousness
  - I am always prepared.
  - I pay attention to details.
  - I get chores done right away.
  - I like order.
  - I leave my belongings around.
  - I shirk my duties.
  - I often forget to put things back in their proper place.
  - I make a mess of things.
- Extraversion
  - I don't mind being the center of attention.
  - I feel comfortable around people.

  - I start conversations.
  - I am the life of the party.
  - I think a lot before I speak or act.
  - I don't like to draw attention to myself.
  - I have no intention of talking in large crowds.
  - I am quiet around strangers.

- Agreeableness
  - I am interested in people.
  - I sympathize with others' feelings.
  - I make people feel at ease.
  - I take time out for others.
  - I insult people.
  - I feel little concern for others.
  - I am not really interested in how people feel.
  - I am not interested in other people's problems.

- Neuroticism
  - I change my mood a lot.
  - I get irritated easily.
  - I often feel blue.
  - I get stressed out easily.
  - I worry about things.
  - I get upset easily.
  - I am relaxed most of the time.
  - It is not easy to make me feel blue.

As far as the method of computation is concerned, each question is assigned a value of accuracy for the user by the user himself (this method of describing one's personality is called self-assessment). The possible values range from 0 to 6 but can be changed in the code very easily. Questions marked in red have inverted values, e.g. if a red question is assigned a value of 5 by the user, we use the value 1 (and vice versa). The black questions correspond with the factor (e.g. "I change my mood a lot." corresponds with Neuroticism), the red ones contradict it (e.g. "I am relaxed most of the time." contradicts neurotic behaviour). A percentage is then calculated and produces a number of 0 to 1 for each OCEAN factor.

5.4.3 Interstep

The aforementioned OCEAN factors have to be presented to the user as a certain kind of reward for participating in this exploratory investigation. This presentation is done in the interstep page. An exemplary result is shown in figure 5.3. It is important not to stop here and allow the application to access user's data through the Facebook profile. This is the purpose of the `Facebook` button.

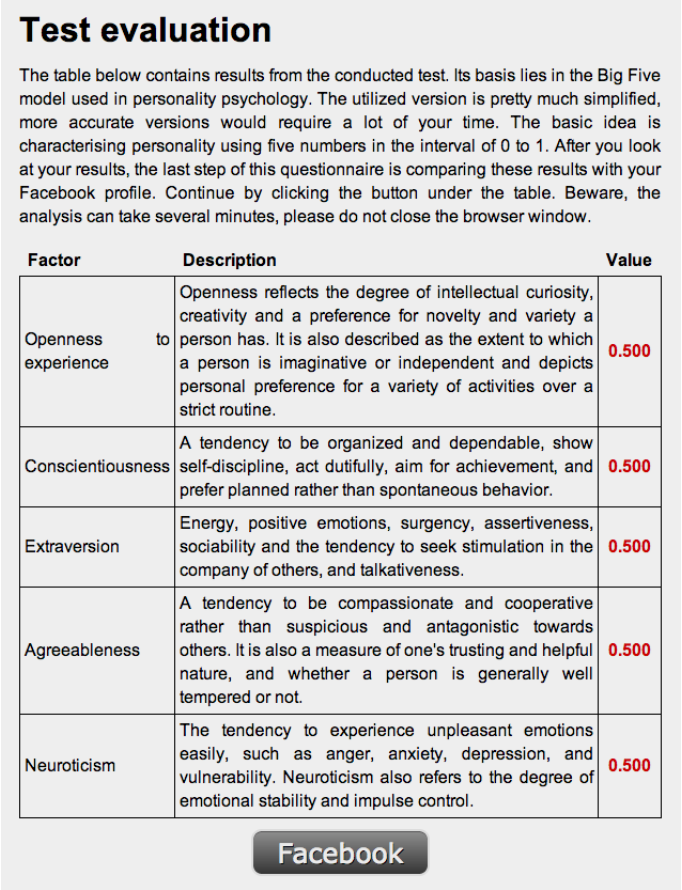


Figure 5.3: The interstep screen showing OCEAN results.

5.4.4 Facebook results

As all the other pages are just administrative support for this page, the Facebook results page is the core of the application. The user is asked for permission to access his or her profile and then the friends of the given profile are sorted by the amount of supposed trust the user puts in them (as seen in figure 5.4). The user is also given a brief explanation on the methodology of trust computation and is given the opportunity to mark the people he or she thinks are in the wrong place in the list. This might prove as an important benchark of this model's success.

Name	dof 1	rle 1	cop 1	atc 2	aoo 4	evp 2	aoc 3	roc 4	sum	Wrong!
	1.000	1.000	0.103	0.725	0.994	1.000	1.000	0.705	0.853	☐
	1.000	0.000	0.782	0.141	0.999	1.000	0.792	0.474	0.685	☐
	1.000	0.000	0.103	0.725	0.993	0.000	1.000	0.685	0.681	☐
	0.404	0.000	0.090	0.725	0.997	1.000	1.000	0.218	0.656	☐
	1.000	0.000	0.269	0.887	0.988	0.000	0.264	0.964	0.647	☐
	1.000	0.000	0.615	0.010	0.997	1.000	1.000	0.206	0.636	☐
	1.000	1.000	0.564	0.082	0.968	0.000	0.264	0.884	0.607	☐
	1.000	0.000	0.115	0.597	0.999	0.000	1.000	0.390	0.604	☐
	1.000	0.000	0.462	0.000	0.986	1.000	1.000	0.056	0.590	☐
	1.000	0.000	0.090	0.725	0.995	1.000	0.264	0.313	0.587	☐
	1.000	0.000	0.038	0.000	0.958	0.000	0.528	0.989	0.578	☐
	1.000	0.000	0.103	0.725	0.990	1.000	0.528	0.076	0.578	☐
	1.000	0.000	0.577	0.010	0.966	0.000	0.264	0.994	0.568	☐
	0.404	0.000	0.090	0.725	0.994	1.000	0.264	0.298	0.550	☐

Figure 5.4: The table summarizing the Facebook analysis results.

5.4.5 Questionnaire conclusion

After marking the people the user deems placed in the wrong order in the list and hitting the `Send` button, the questionnaire is finished. The final page thanks the users for participation and in the background stores the acquired values into the prepared database where they are processed. Their final interpretation is described in detail in chapter 7.

Chapter 6

Implementation

Although not a key part of the thesis, it was still necessary to implement the model in a programming language to be able to evaluate users' participation. This chapter describes notable insights in the matter of Facebook data mining and also Facebook application development. Gathered findings can be used as a guide for developers interested in exploring the studied matter further and also to build upon the outlined framework of trust-related or psychological development. It is also necessary to mention (at least in a brief manner) the equipment offered by the Faculty of Information Technologies at Brno University of Technology [oT08]. Implementation of individual contexts is described in 3.2 for clearer continuity.

6.1 Development and working environment

There are several ways how to create an application working on the Facebook platform. As a very extensive and detailed explanation is given at the Facebook Developer website [Fac08b], it would be uproductive to describe this process in the thesis. Instead focus is given to the pecularities and traits of this particular application instance.

All Facebook applications need hosting. This option is offered by Facebook itself, e.g. described in the article [Fac08a]. As this is a paid option, university resources were used instead. Faculty of Information Technologies in Brno University of Technologies gives its student an opportunity to use the server `merlin.fit.vutbr.cz`. Its parameters and properties are described in the table 6.1. Usage of this server is of course very limited and cannot be used for commercial purposes because of legal issues. Nevertheless, the given capacity matches the intended scale of the model completely and makes it possible to use hosting for academic purposes.

Server hardware information		**Software information**
Address	`merlin.fit.vutbr.cz`	PHP 5.3.3 (cli)
CPU	AMD Quad-Core Opteron 2387 2.8 GHz	MySQL 14.14 (Distribution 5.1.73)
RAM	33022208 kB	phpMyAdmin 3.4.5
IP	147.229.176.19	Sublime Text 2.0.2 (Build 2221)

Table 6.1: Application environment for the Facebook application.

As for the chosen development path, this can already be assessed based on the given

software versions. Again, developers can choose from a wide variety of programming languages. Prepared SDKs include iOS, Android, JavaScript, PHP and Unity. Several of these options had been tried out and PHP was chosen in the end. The following list relies heavily on the source [Fac08c] and enumerates the main reasons which justify choosing this path:

- JavaScript would require the user to turn it on in their browser for this application, making it less likely to get potential respondents,
- PHP utilizes the server resources for advanced FPU unit operation, which helps the computation in context "Regularity of contact" 3.3.3,
- as all the actions are taking place on the server and the user is only given a static page in the end, no additional security measures (as opposed to JavaScript) are needed,
- computation is encapsulated on the server and therefore reduces the effort for browser customizing,
- PHP was chosen for the greatest amount of previous experience with web application programming in this language.

To conclude the description of used environment, table 6.2 contains all the browsers the application was tested on. As had been stated before, all the computation takes place on the server, which leaves little room for browser-specific issues. Nevertheless, for the widest possible range of potential respondents it is crucial to make this application work in as many different settings of environment as possible.

Browser name	**Version**
Google Chrome	34.0.1847.116
Mozilla Firefox	28.0
Safari	7.0.3 (9537.75.14)
Microsoft IE	11.0.9600.16521

Table 6.2: Browsers tested for the current application version.

6.2 Modular structure

One of the greatest flaws of the bachelor thesis [Šve11] was the absence of any division of programming code into modules according to the opponent. Even though the design was object-oriented and functionality encapsulated into Python classes, the code was quite difficult to read and understand. Implementing the model in PHP also had one of the primary goals set to programming in a modular way. The advantages of modular programming are described in many various workings of renowned computer scientists, one of them being the source [GJJB05].

This section describes the division into individual modules (see figure 6.1) and also briefly introduces their functionality. It is not feasible to go into deeper detail in this section, as the source code turned in together with this master's thesis is freely available. Functions are documented directly in the code. No additional libraries (apart from the PHP SDK for Facebook) were used in the creation of the application.

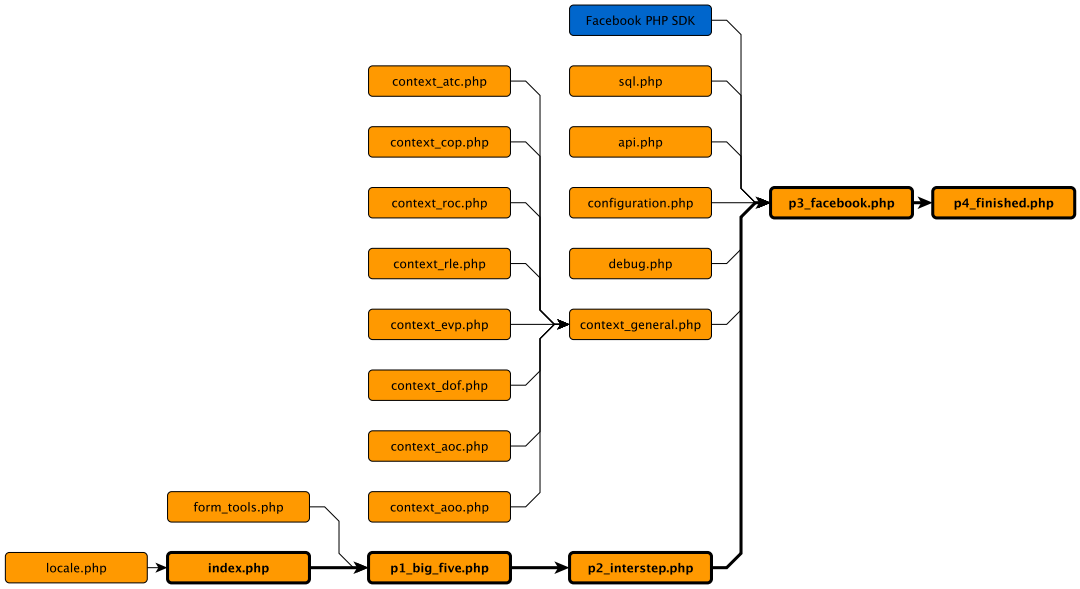


Figure 6.1: Modular structure of the application.

6.2.1 Facebook PHP SDK

This SDK is offered by Facebook Developers to make development of Facebook aaplications easier for the PHP language. Details can be found in source [Fac08c]. The PHP SDK utilizes a native library to access the Graph API and also implements the Login functionality. Its usage is very intuitive and easy to understand. The only thing needed for setup is downloading the SDK directly from Facebook Developers (ca. 430 kB unzipped) and uploading it into the folder which contains the application on the server. This gives the possibility to require the `facebook.php` file via command `require_once('php-sdk/facebook.php');`.

The SDK then offers all the necessary functionality to manage the application, such as login with determined permissions (see 6.4 for details), managing login address, determining the user ID of currently logged user and also API calls, which are used abundantly in the final version of the application. The PHP SDK forms the majority of functionality in the module 6.2.3.

6.2.2 sql.php

The SQL database concept is described in detail in section 6.5. This module includes a function for MySQL database initialization in function `sql_init()`. The application connects to a designated database on the `merlin.fit.vutbr.cz` server, selects the database intended for this purpose and sets proper encoding. In the end it computes the time stamp for the last point in the past when data was valid. This concept is in detail described in the parameter `VALIDITY_INTERVAL` in subsection 6.2.4. Designated functions are then available for inserting (`sql_insert($request, $data)`), deleting (`sql_delete($request)`) and getting information from the database (`sql_get_request($request)`). The module also serves for storing results acquired from the transitional superglobal `$_SESSION`.

6.2.3 api.php

As this is one of the most extensive modules, only a selected number of functions shall be explained. The module needs to be initialized once at the beginning of usage (`api_init()`). Functions `api_std_limited($request_string)` and `api_std_nonlimited($request_string)` utilize the standard Open Graph API described in detail in section 6.3. The limitation depends on whether or not we want to limit the request on API with a specific point in the past where we are no longer interested in the result (parameter `ANALYSIS_SINCE`). The same concept is used in functions `api_batch_limited($batch_array)` and `api_batch_nonlimited($batch_array)`, which utilize the batch Open Grap API. The function `api_fql($request))` does not use the time limitation concept, as its usage is very narrow in the final application.

6.2.4 configuration.php

Module `configuration.php` brings all possible configuration data and constants into one file. It is very convenient for experiments with the parameters and also as a storing point for all the accesses. Potential users should be aware of the fact that exposing this file is a serious security risk, as it contains access credentials to the `merlin` database and to the application on Facebook. This file may by no means be accessible to anyone but the application developer. The table 6.3 briefly introduces the meaning of parameters.

6.2.5 debug.php

A very simple and short module which implements the functionality for debugging messages. According to the debugging level set in file `configuration.php` the function `debug_print($debug_object, $level)` either prints the text in HTML or does not. An example of usage would be the PHP code `debug_print(''Friend ID: '' . $friendID, 5);` where the number 5 is assessed against the current `DEBUG_LEVEL` value. An additional function was implemented to be able to print nested arrays with their hierarchy. Printable objects in the array can either be in the `String` or the `DateTime` class.

6.2.6 context_general.php

Module `context_general.php` unifies access to individual contexts of trust being computed. As such, the module provides only limited functionality. Context computations are called one by one and results stored into an array. The function `microtime(true);` is also used for benchmarking purposes to measure time which individual contexts take to compute. The contexts are then aggregated into a single value for each friend of the user. All these actions happen in function `context_general_compute()`. Contexts and their summary are then printed in a synoptical HTML table in function `context_general_print_contexts($contexts)`.

6.2.7 context_XXX.php

As the computation process for each context is already described in 3.3, very little can be said about these modules themselves. The letters `XXX` are substituted with one of the 8 abbreviations of contexts used in the model. Every module has a unified entry point `context_XXX_compute($friends)`.

6.2.8 index.php

This is the entry point into the application. Its primary tasks are very simple - encoding is set to UTF-8 in both PHP and in HTML as meta charset. PHP SDK is required. This module sets the language for both the following questionnaire and also for the description of Facebook analysis results into the `$_SESSION` variable, which also serves for passing values later on.

6.2.9 p1_big_five.php

The questionnaire module. It is important to mention there are two ways to describe the implementation of the questionnaire. This is the technical one, you can find the semantical one in section 5.4. As for the technical side of the problem, the data collection is fairly easy. In each of the five factors of the OCEAN model, 8 questions are presented in the form of sliders (HTML text is acquired from the module `form_tools.php`). A notable characteristic in this case is the use of the `<intput type='range'/>` tag from HTML 5, which is displayed differently in virtually every single browser used for testing. Therefore it was necessary to use CSS styles and adjust the looks.

After presenting the user with sliders for 8 questions, clicking the button `Next` executes the evaluation of the current factor and stores it into an internal variable in `$_SESSION` called `current`. The `$_POST` data transfer HTTP method is used. The page is then refreshed from PHP and presents the next factor or (in case of the last Neurotocism factor) the interstep page described in the next module.

6.2.10 p2_interstep.php

The module for interstep only summarizes the values acquired in the OCEAN sliders collection and presents them to the user. The main technical meaning of this module is to acquire privileges to the user account on Facebook to be able to start the analysis itself. More details on the motivation can again be found in section 5.4.

6.2.11 p3_facebook.php

The Facebook module is the main part which utilizes the main `context_general.php` module. The SQL database for storing results is initialized, API is initialized as well while checking whether the user has given the application sufficient access rights. It then conducts the analysis (which may take several minutes) and presents the table with results and also the explanation of the basic mathematical background to the user. The resulting statistics are computed (taking into account several views, for details see chapter 7) and stored into the superglobal variable `$_SESSION`.

6.2.12 p4_finished.php

Module `p4_finished.php` thanks the users for participation and displays the final text of the questionnaire. Simultaneously the module `sql.php` stores the results together into the pre-designed SQL database format and allows us to look for correlation in the values.

6.2.13 form_tools.php

Used by module `p1_big_five.php`, the form tools offer basic functionality for the Big Five questionnaire. It only contains the functions to print the slider piece of HTML code, collect the values and then display the result to the user (also in form of HTML output). A notable feature here is the possibility to easily change the slider scale in the code using constants and determine how many steps are selectable for the user.

6.2.14 locale.php

Only one function is implemented: `locale('id')`. This function returns the string of the determined ID of text based on the language in the `$_SESSION` variable. As this variable is superglobal, function `locale` can be used throughout modules and its purpose is to make it possible to implement the application in more languages. In this case, Czech and English were used. The PHP file then contains a two-dimensional array composing of texts in the two (or, with very little additional implementation effort, more) languages and provides the possibility to conduct localization all in one place.

6.3 Open Graph utilization

A short notice should also be given to the Open Graph concept. New developers can find detailed introduction into this topic in source [Fac08b]. To keep the explanation brief and simple, Open Graph is an Application Programming Interface (API) for Facebook data. Its input is a string created according to the set hierarchy (i.e. `/me/friends` extracts a list of friends of the currently logged user). Its output is a JSON object in an anticipated format containing the requested information. The following subsections describe some issues and functionality a new developer may encounter on the way.

6.3.1 Paging

Probably the most unpredictable feature in the whole Open Graph API, the reasons for this behaviour are given in a How-To article written by Jeff Bowen [Bow11]. If the requested amount of data exceeds the amount deliverable by one JSON object, it is divided into several "pages". They are in the form of an additional JSON object called `paging`. This object gives addresses for additional results in both directions and the developer can then browse the results like a book. So far this concept seems like a beneficial idea and its existence is fairly justifiable.

The disadvantage of the Facebook implementation is the fact that we cannot anticipate the paging in advance and even the same request may give us various results in terms of paging in a certain span of time. This phenomenon is described in the aforementioned article in figure 6.2. Step 2 shows Facebook removing results not visible to the viewer. As this is an internal step conducted by Facebook itself, there is no way one can anticipate the result. The solution, while bothersome, is fairly easy to implement and involves expecting pagination in any obtained data.

6.3.2 Batch concept

Every Facebook application which exceeds only basic functionality is bound to need a higher number of API requests. Open Graph therefore offers the possibility to bundle these

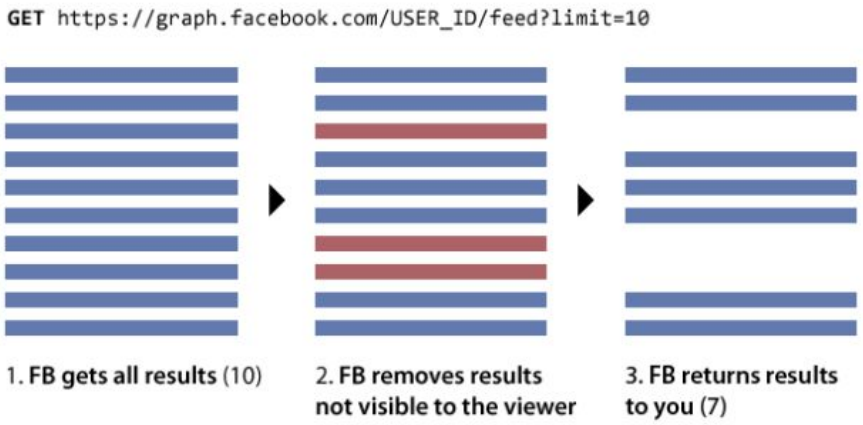


Figure 6.2: Reason for various paging results over time.

requests into a single hierarchical JSON object in order to reduce the overhead needed for maintaining so many connections. Facebook Developers describe this concept in a detailed article in source [Bat11]. Utilizing this concept leads to significant rise in performance (implementing batch requests decreased the runtime of the finished Facebook application by more than 400%). It is important to note that Facebook only allows up to 50 requests in one batch.

6.3.3 FQL

The Facebook Query Language is an obsolete subset of the Open Graph API which will no longer be supported in the upcoming versions [FQL11]. It is designed to utilize the concepts introduced in the SQL language and for developers familiar with this language it may seem more intuitive than the Open Graph standard requests. The language, however, does not provide any additional functionality or benefits and just serves as a bridge to offer a more familiar environment. In rare cases the queries can be simpler to write in this language, as shown in context Contact preferences (see 3.3.6).

6.4 Permissions concept

When a user logs into the application via Facebook Login, the application can access a subset of that user's data stored on Facebook. Permissions are a way of asking someone if the application can access that data. They are represented by strings that are passed along with a login request or an API call. When a user logs into the application, a dialog will be shown to the user to request the permissions that the developer requests in the application [Fac11]. The following is a list of permissions needed for this particular application:

- read_stream,

- user_photos,
- friends_status,
- user_groups,
- user_events,
- friends_events,
- read_mailbox.

6.5 Database caching

As the time consumed for a profile analysis can reach up to several minutes, it is not very practical to develop the application under these circumstances. It was therefore necessary to ensure a kind of caching for the requests on the Facebook API. The solution to this issue is a simple MySQL database underlying the whole application. Each time data is requested in the application, the database is checked. If the entry is already present in the requested timespan (see configuration parameters in table 6.3), the application does not issue the request. If the current data is missing in the database, API is utilized and the result simultaneously stored in the database for future use.

From the technical point of view, it was necessary to edit the JSON objects in order to store them in the SQL database in plaintext. The objects are serialized and unserialized using functions `serialize` and `unserialize` in PHP. Functions `base64_encode` and `base64_decode` are then used to solve the issues with SQL database input, where some symbols interfered with the designated ones in the SQL queries.

Taking into consideration the privacy issues for the users taking part in the resulting questionnaire, the database caching was turned off in the final version of the application. It can be turned on for future development of additional functionality.

Parameter	Meaning	Notable content limitation
APP_ID	The identificator under which the application is registered with Facebook.	The ID for the "Trust in social networks" application is 212701342136307 for future reference.
APP_SECRET	The application secret serves as a kind of password that only allows developers who registered the application to actually use it in their code.	
PERMISSIONS	Each user of the application is asked for access to some information on his or her profile. This parameter defines which permissions are required.	String with each individual permission separated by comma, see 6.4.
DEBUG_LEVEL	Sets how much debugging text the user shall see. Standard users should have this value set to 1.	See 6.2.5.
DB_ADDRESS	The address of SQL database the application uses, in this case the MySQL database on `merlin`.	
DB_USER	Username for the database.	
DB_PASSWORD	Password for the database.	
USE_CACHING	Whether or not caching in the database should be used.	Acceptable values are `TRUE` or `FALSE`. See 6.5.
VALIDITY_INTERVAL	Determines the timespan for cached requests in the SQL database which are still valid.	See [Gro10].
ANALYSIS_SINCE	Date since which the analysis is conducted. Excessive values (i.e. one year) may cause unnecessary delays and inadequate server load.	String in format 'YYYY-MM-DD'.
PRIORITY_VECTOR	Determines weights for individual contexts of trust.	See 3.5 for explanation.
MAX_BATCH_REQUESTS	Sets the maximum number of Open Graph requests in one batch.	This parameter can be set up to 50. See 6.3.2.
TOP_PERCENT	Sets the amount of top users taken into consideration for statistics.	The value must be in the interval between 0 and 1.
HOMEPAGE	The homepage used for links.	Set to the `merlin` link of the application.

Table 6.3: Brief meaning of parameters defined for the application.

Chapter 7

Experiment

The final chapter of the thesis summarizes and explains the results acquired through the conducted exploratory investigation. The origin of the data is explained, then several metrics and views are used to observe this data and find correlation to statements made in the previous chapters. Both the general and the specific approach are used. OCEAN factors are observed with no connection to other data and the same is done with the trust values. After this one-dimensional observation, these two quantities are observed together using methods to find correlation.

7.1 Data collection

Data used in this experiment was acquired in the course of three days from 21.5.2014 till 23.5.2014 using the questionnaire described in 5.4. The invitation to try the application was spread through Facebook using the sharing method and in the end 47 users participated in the exploratory investigation. The data set was collected using a MySQL database on the `merlin` server and collected using the tool phpMyAdmin installed on the same machine. The export was done into the Microsoft Excel format and then further processed using the Excel tools together with GIMP 2.8 to enhance the quality of resulting images. One interesting fact to mention is the amount of data was too overwhelming for Microsoft Excel to process and a simple PHP application was therefore used for the view 7.6.

7.2 Execution time and efficiency

The first observed quantity in the model was the time of execution for individual context computation. As users are used to web applications running in almost real time, it was difficult to convince them to persevere on the website and wait until the analysis is done (and some of them did indeed leave the page before it was ready). The two following figures explain the problem in a more understandable way.

Figure 7.1 shows a chart of average time consumed by each context for the 47 users. The time was measured using the PHP function `microtime` and indicates the time between the start of the specific context computation and the moment the computation was finished and stored into a variable. This was done for all 8 contexts and for all users and average then computed for all the user data.

Two phenomena can be observed here. One of them is the duration of the `atc` context, Allegiance to communities. It is quite natural to be the highest, since the context needs all

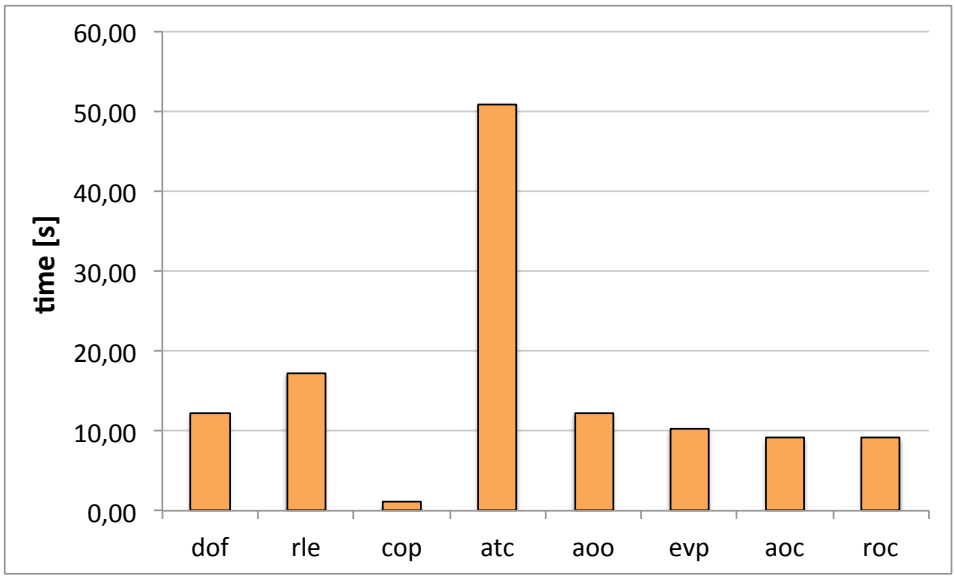


Figure 7.1: Average time spent on computing contexts.

groups the user belongs to and also all groups his or her friends belong to. On the other hand, the context `cop`, Contact preferences, took only a little over one second to compute and indicates nicely why the FQL method, while deprecated in the last versions of API, was utilized. The short time is due to only one query which was necessary to be sent.

Figure 7.2 puts the previous data into perspective with the priority given to individual contexts. Priority gives us an overview on how much the context contributes to the final result. Context `atc` is clearly an exception and due to its ratio of contribution to time, it is advisable to omit it in the upcoming versions of the model. In need of optimization the `rle`, Real-life experience, can also be omitted, as it consumes the most time out of the least significant contexts.

7.3 OCEAN characteristics

The OCEAN characteristics were estimated based on the implemented questionnaire. The average values were then computed for each individual factor in order to figure out what the distribution of these characteristics looks like in the given data set. Another reason to compute this average is its usage in the later analysis described in section 7.6.

Figure 7.3 shows the result. As can be seen in the chart, two factors show significant deviation from the average value of 0.5: Openness to experience and Conscientiousness. This phenomenon can be explained in a satisfiable manner by the sources [Ave25] and [Boi08], which suggest that younger adults tend to achieve higher scores in the factors O and C. The values have a tendency to deteriorate in the course of time. As the age of respondents lies within the range of 15 to 32 years, the above mentioned finding does apply to them and explains the deviation.

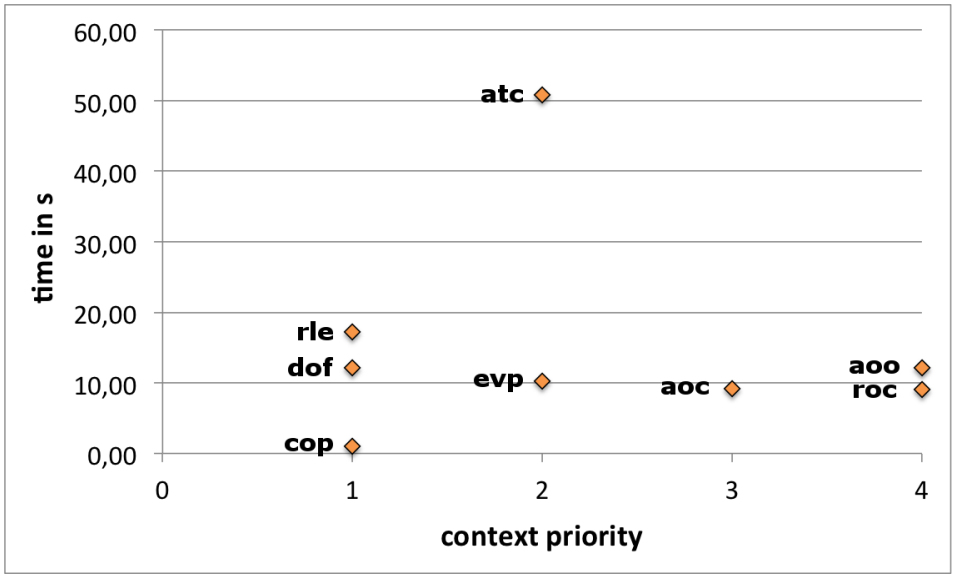


Figure 7.2: Context computation time in correlation with priority.

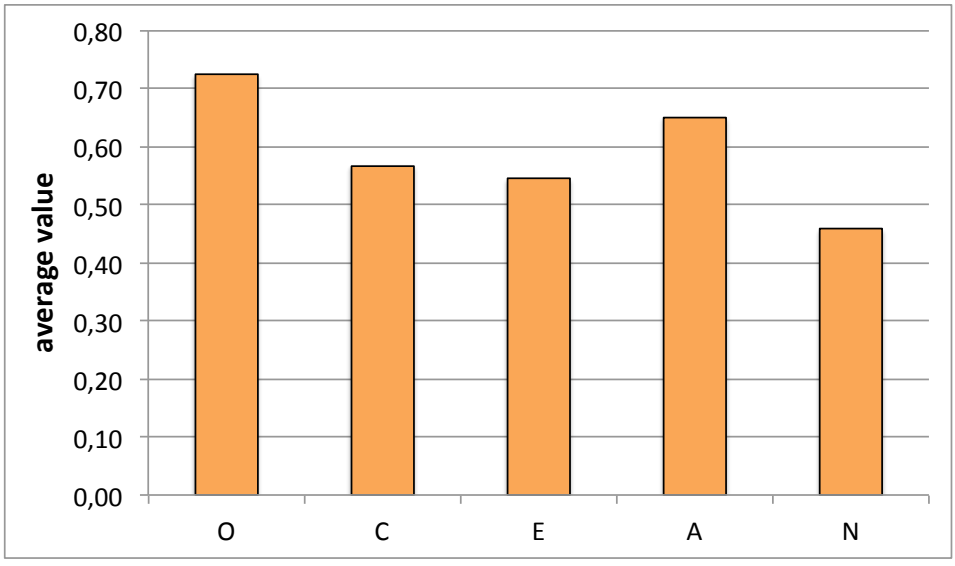


Figure 7.3: Average OCEAN factors of the respondents.

7.4 General characteristics of trust

This section analyzes resulting trust as such and does not take into account the personality at all. The primary goal of this view is to optimize this model in pursue of greater factual accuracy. Several figures shall be presented, each explaining a different trait of the model. In case of potential defects an explanation shall be given as to how to fix the model in the most convenient way.

7.4.1 Average trust

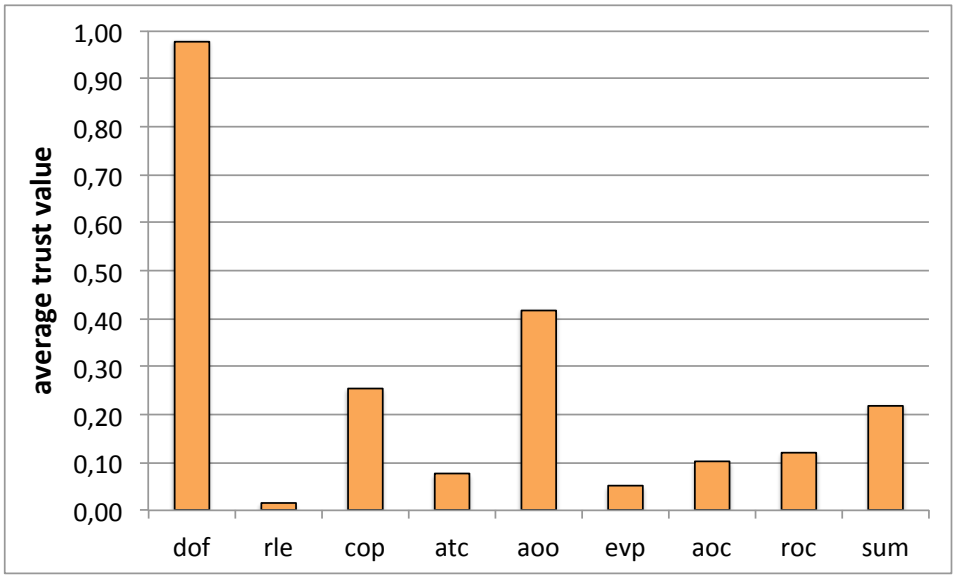


Figure 7.4: Average trust of the 47 respondents.

Figure 7.4 shows the average trust computed for all the 47 participants of the exploratory investigation. It shows the greatest fallacy of the model which was possible to discover only using real-world data from users and which should definitely be taken into consideration in further development. In an ideal model, the average values should be distributed around the same value for all contexts. The problem is caused by the priority vector, which is primarily designed with the thought that the average value of all contexts is the same. If it is not (such as in this case), the model shall be biased by the priorities. The results are still usable, yet another solution of the presented problem should definitely be implemented.

One suggested approach is to sort the users for each individual context based on their value and then distribute them evenly in the interval. As an example, we take three users with the `dof` value of 0.85, 0.91 and 1.00. We would sort them by trust and assign proportional values, in this case 0.25, 0.50 and 0.75. The average would then be 0.5 and using more contexts in the priority vector would be safe.

7.4.2 Average trust of top users

Another interesting aspect to compare is how the average trust is formed in the top 15% of friends of each individual user. The figure of 15% was taken over from the bachelor

thesis [Šve11] and indicates the area in which the model was most successful in its previous implementation. The results of the comparison can be seen in figure 7.5.

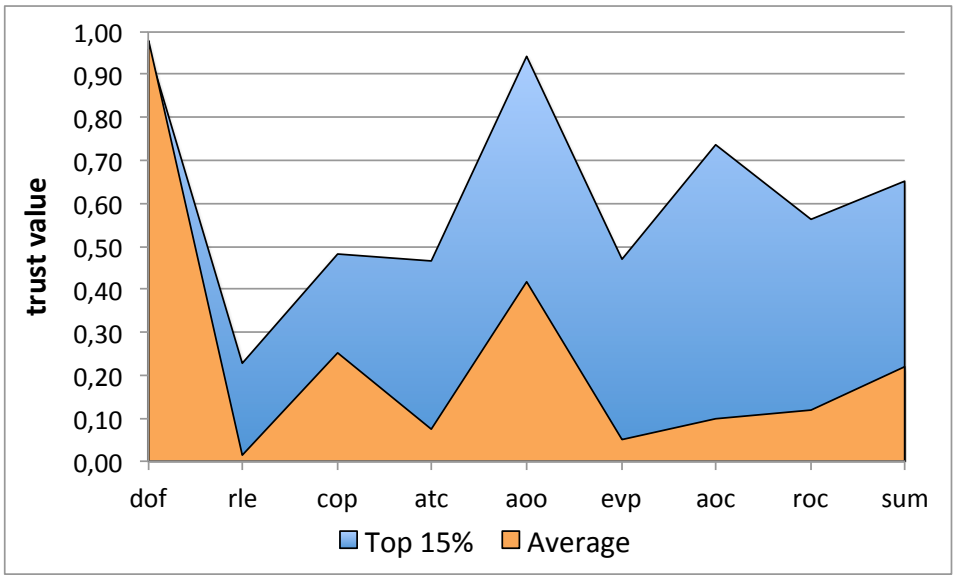


Figure 7.5: Trust of the top 15% users compared to the average.

Several facts should be mentioned here. Presumably, for the 15% of our most "elite" friends the values should be higher in all contexts. This, however, is not the case for context `dof`, which is lower for the top users. This result indicates that it does not in fact matter so much how long we know people in order to form trusting bonds with them. Another interesting spike can be seen in the `aoc` context. For the top users this context seems to be even higher than expected and suggests assigning this context a higher priority.

The third and last notable difference in the two sets of values is in the `rle` context. There is a steep difference between the average and the top, suggesting we indeed tend to have much more common photographs with our top trusted friends, and this fact revealed itself despite its lower priority value.

7.4.3 Absolute deviation

The last insight on general trust involves the absolute deviation and is depicted in figure 7.6. After computing the average and assigning it as a central value into the absolute deviation formula, the results seem to correlate quite nicely with the set priority vector (apart from contexts `cop` and `roc`, their score is a bit deformed). As much as this would indicate an efficient solution to priority vector stabilization, it is important to keep in mind the findings from subsection 7.4.1. As the current priority vector was assigned empirically, there is high probability that higher priorities were assigned to more diversified values. The suggested solution would bring absolute deviation of all contexts to the same level.

7.5 Inaccuracy distribution

Users were given the opportunity to state whether the computed trust fulfilled their expectations of the model. Checkboxes next to each name allowed them to mark people who had

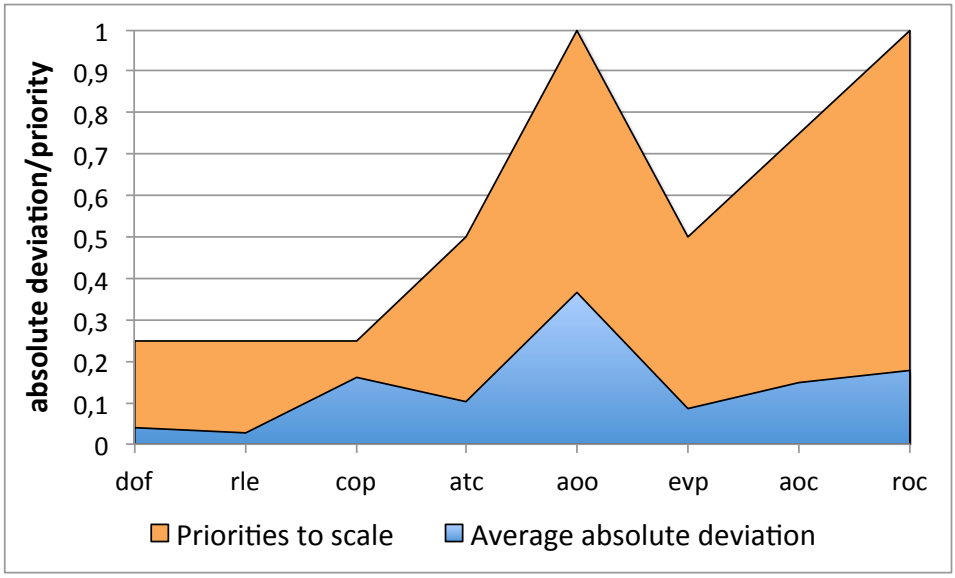


Figure 7.6: Absolute deviation of trust compared to priorities.

extra or not enough trust from their point of view. The average amount of mistakes per user is an average of 6.62 (for scale, the average number of friends for one user was 220.77). This value was largely affected by one thorough user who selected 67 friends. Median value for this set was actually only 3.

Figure 7.7 shows the distribution of friends in the list ordered by the summarized trust values. The chart shows a larger amount of people selecting number one (as it is quite difficult to find the person who the user trusts most of them all) and then a spike after position 10. This can be seen as the weak spot of the model and should be improved in future versions. The steadily decreasing number of errors towards the end can be ascribed to users not being interested in their less trustworthy friends, which on the other hand means that top users were picked with sufficient accuracy.

7.6 OCEAN versus trust

This section is the core product of the master's thesis. As far as the previous chapters are concerned, they just serve as support for finding correlation between trust in the computational sense and the personality of users, overlapping from the area of artificial intelligence into user modelling.

7.6.1 Used methodology

Finding correlation in two quantities is not an easy task. The method used to find any connection between the Big Five model and computational trust was empirically chosen after several futile attempts to visualize the data. The final dataset was chosen to be the absolute deviations from the average values for both trust and the OCEAN factors. Therefore we have 5 factors and 8 contexts to consider, giving us 40 possibilities altogether with several possible outcomes for each case. This number of possible outcomes calls for computational processing.

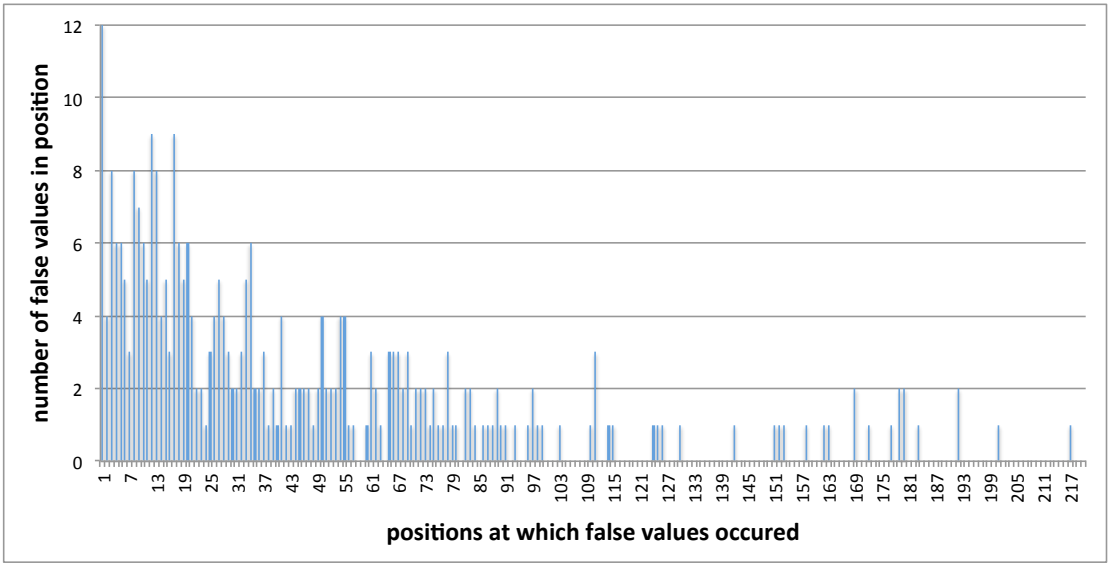


Figure 7.7: Distribution of errors in list positions.

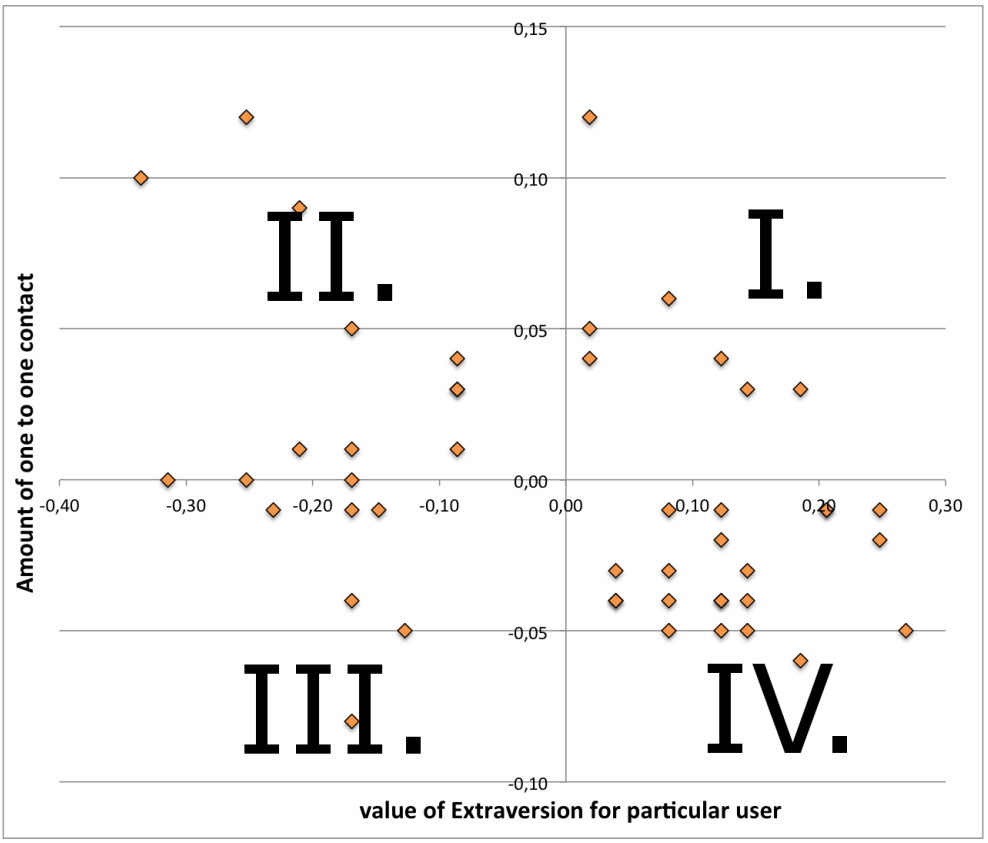


Figure 7.8: The method of finding correlation using quadrants.

A basic example of what we are looking for is given in figure 7.8 (please consider this only as an example, it does not correspond to any real data). The x axis represents the chosen OCEAN factor, the y axis represents the average trust for the given user in the given context. In this example, it is the `aoo` context and Extraversion. Each orange point in the scatter chart represents a single user. The plane is divided into 4 Cartesian quadrants. These quadrants point to a certain relationship of the two charted quantities only under certain circumstances:

- users whose values in the x or y axis reach 0 are excluded,
- any relationship of two neighbouring quadrants points to a bias to only one quantity and is therefore useless,
- the delta has to be of significant importance, not just a random swing,
- only individual quadrant swings or swings of opposite quadrants shall be considered.

With the above rules in mind, the analysis of the given figure 7.8 gets somehow clearer. We can now count the users belonging to individual quadrants and compare their numbers:

- quadrant `I`: 7,
- quadrant `II`: 9,
- quadrant `III`: 6,
- quadrant `IV`: 18.

The picture we get now is that the `IV`. quadrant is somehow significant. It fulfills all the set requirements and therefore we could assume that respondents with higher values of Extraversion tend to trust less in personal messages. This methodology was applied to all the data in the following section 7.6.2.

7.6.2 Inferred conclusions

As the possibilities are numerous for the particular task of finding correlation, a computational approach had to be taken. A simple PHP application produces an HTML table with numbers of users for each quadrant and for each combination. Each table cell for the OCEAN/trust combination contains a miniature model of the four quadrants (for reference plese see 7.8). The table is shown in figure 7.1. Significant values are marked in red and then commented further.

As the results are quite difficult to put into words comprehensively, another table is necessary. The table 7.2 shows the trust context, the OCEAN factor, correlation expressed in words and its possible explanation. It is important to take into consideration the possibility of the data bias, as 47 users do not provide such a rich data set. Only combinations containing a swing of 5 or more from the mean of the given combination were inspected (and in the case of the DOF context some were discarded).

	O		C		E		A		N	
Durability of friendship	12	15	12	11	10	17	9	18	12	14
	6	4	4	6	6	4	5	5	4	6
Real-life experience	9	6	5	9	6	9	6	9	9	6
	11	9	8	9	7	13	8	12	8	11
Contact preferences	10	10	6	12	10	10	10	10	11	9
	14	11	11	12	7	18	9	16	12	12
Allegiance to communities	8	9	5	10	10	7	6	11	9	8
	15	12	12	13	6	21	12	15	13	13
One-to-one contact	9	11	8	11	7	13	6	14	11	9
	15	12	11	13	12	15	14	13	12	14
Events participation	9	9	6	9	3	15	6	12	9	8
	11	9	9	10	13	7	10	10	11	9
Amount of contact	13	11	9	14	8	16	8	16	11	13
	10	10	10	7	11	9	11	9	10	10
Regularity of contact	12	12	9	13	7	17	8	16	10	14
	9	11	8	10	10	10	11	9	12	8

Table 7.1: Distribution of users in quadrants in OCEAN vs. trust.

7.6.3 Summary of correlation

The expected correlation described in table 7.3 (for more comfortable comparison displayed once again) was only partially fulfilled. The result could perhaps have been anticipated by an experienced sociologist or a psychologist with the area of interest in social networks. It is interesting to compare the table 7.4 to the estimated one, as the values actually meet in the contexts with estimated correlation to Extraversion and sometimes Agreeableness.

The main fact this table shows is that trust in social networks is largely dependent on extraversion (represented by the same Big Five factor) and on the desire to be liked (represented by the Big Five factor Agreeableness). One additional dependency was discovered in the Big Five Conscientiousness factor, although this exception might not have anything to do with social networks. It might be connected to the dilligent administration of one's groups on Facebook.

		Statement	Possible explanation
DOF	A	More agreeable people tend to trust people they know for a longer time.	Agreeable people do not like conflicts and if they stayed friends with someone for a longer time, they tend to believe there will be no conflicts with them.
COP	E	More extraverted people tend to trust people they have not many common friends with.	Extraverted people trust also people outside of their "stable" circle of friends.
ATC	C	Few less conscientous people tend to trust people with common groups.	The explanation could lie in the fact that less conscientious people do not administer their groups very well.
ATC	E	More extraverted people tend not to trust people with common groups.	This is a case similar to the COP context. Extraverted people also trust people outside of their stable groups.
ATC	A	Few less agreeable people tend to trust people with common groups.	Agreeable people may find assurance mechanisms for trust in common groups. Therefore groups serve as a means of maintaining trust outside of their own personality.
AOO	A	Few less agreeable people tend to trust people with whom they maintain one-to-one communication.	Again, agreeable people need assurance and harmony which leads them to think that if their friends do not communicate with them one-to-one frequently, they are not trustworthy.
EVP	E	Extraversion is in direct correlation with event participation.	The result is not surprising. Extraverted people tend to trust people they attend events with.
AOC	E	Extraverted people tend to trust people they have more contact with.	Amount of contact seems to be crucial for both extraverted and agreeable people in order to trust someone.
AOC	A	Agreeable people tend to trust people they have more contact with.	The explanation is similar to the case of AOC and E.
ROC	E	Extraverted people tend to trust people they have regular contact with.	One of the nicest correlations in the last 4 items in the table. Not only amount of contact, but also its regularity is important to extraverted and agreeable people.
ROC	A	Agreeable people tend to trust people they have regular contact with.	The explanation is similar to the case of ROC and E.

Table 7.2: Overview of trust and expected OCEAN relevance.

Context	Relevant factors				
	O	**C**	**E**	**A**	**N**
Durability of friendship	0	1	1	2	3
Real-life experience	2	0	3	1	0
Contact preferences	2	0	2	1	0
Allegiance to communities	2	1	1	3	0
One-to-one contact	0	0	1	1	1
Events participation	2	0	3	1	0
Amount of contact	2	0	3	1	0
Regularity of contact	3	1	2	1	1

Table 7.3: Overview of trust and expected OCEAN correlation.

Context	Relevant factors				
	O	**C**	**E**	**A**	**N**
Durability of friendship				X	
Real-life experience					
Contact preferences			X		
Allegiance to communities		X	X	X	
One-to-one contact				X	
Events participation			X		
Amount of contact			X	X	
Regularity of contact			X	X	

Table 7.4: Real trust and OCEAN correlation for comparison.

Chapter 8

Conlusion

This master's thesis follows up on the bachelor thesis [Šve11] presented at the UMAP 2013 conference [UMA31]. It explains basic terminology for trust and reputation in both information technology and sociology. This terminology recapitulation is followed by the description of an enhanced multi-context trust model. Explanation is given as to which remarks from the UMAP conference were taken into consideration and incorporated into the model.

The most valuable contribution of the master's thesis is the analysis of the field of personality psychology and finding its relevance to trust. Basic terminology for this field is explained and then used to find a metric for comparing users who participate in the conducted survey in the implemented application. Character is chosen to be the metric and the Big Five questionnaire is chosen to be presented to subjects to grasp character in a more machine-readable way. The thesis contains the description of the Big Five model and its expected relevance to the presented trust model.

An exploratory investigation is then conducted on Facebook users utilizing a questionnaire to assess the Big Five factors and find correlation with trust. Two outcomes are worth noting. One outcome brings a new view on the priority vector and suggests multiple further enhancements to the model which can significantly improve its accuracy. The other outcome is the found correlation between trust and OCEAN factors. The conducted exploratory investigation suggests a strong relationship between each user's Openness to experience together with Agreeableness and almost every context of trust contained in the model.

The acquired insights can well be utilized in social network analysis and building trust models for personalized user experience. Although these fields may encounter some barriers regarding bridging the humanities and science into a cooperating relationship, it seems rather difficult, if not directly impossible to explore greater depths of user experience without their mutual cooperation.

Bibliography

[Ave25] Research reports on science from Michigan State University provide new insights [online]. http://www.highbeam.com/doc/1G1-284045791.html, 2012-04-04 [quoted 2014-05-25].

[Bak09] Stephen Baker. *Numerati*. Computer Press Brno, Brno, 2009. ISBN 978-80-251-2444-4.

[Bar54] J. A. Barnes. Class and committees in a Norwegian island parish. In *Human Relations 7*, pages 39–58. London School of Economics, University of London, England, 1954.

[Bat11] Making multiple API requests [online]. https://developers.facebook.com/docs/graph-api/making-multiple-requests/, 2014-05-02 [quoted 2014-05-11].

[Boe09] Big Five minitest [online]. http://webspace.ship.edu/cgboer/bigfiveminitest.html, 2012-07-08 [quoted 2014-01-09].

[Boi08] Sean Boileau. *Openness to Experience, Agreeableness, and gay male intimate partner preference across racial lines*. PhD thesis, Arizona State University, Arizona, 2008.

[Bow11] Jeff Bowen. Paging with the Graph API and FQL [online]. https://developers.facebook.com/blog/post/478/, 2011-03-10 [quoted 2014-05-11].

[CBB11] Gianvittorio Caprara, Claudio Barbaranelli, and Laura Borgogni. *BFQ. Big Five questionnaire*. Giunti Organizzazioni Speciali, Firenze, 2011. ISBN 9788809400672.

[Dun98] Robin Dunbar. *Grooming, Gossip, and the Evolution of Language*. Harvard University Press, 1998. ISBN 0674363361.

[Eba02] eBay [online]. http://www.ebay.com/, 2011-05-01 [quoted 2011-05-02].

[Fac08a] Facebook + Heroku [online]. https://developers.facebook.com/blog/post/558/, 2011-09-15 [quoted 2014-05-08].

[Fac08b] Getting started with custom stories [online]. https://developers.facebook.com/docs/opengraph/getting-started/, 2013-10-25 [quoted 2014-05-08].

[Fac08c] Facebook SDK for PHP [online]. https://developers.facebook.com/docs/reference/php/4.0.0, 2013-10-28 [quoted 2014-05-08].

[Fac11] Permissions with Facebook login [online]. https://developers.facebook.com/docs/facebook-login/permissions/v2.0, 2014-05-01 [quoted 2014-05-11].

[FQL11] Facebook Query Language overview [online]. https://developers.facebook.com/docs/technical-guides/fql/, 2014-05-08 [quoted 2014-05-11].

[Gid90] Anthony Giddens. *The Consequences of Modernity*. Polity Press, Oxford, UK, 1990. ISBN 0745607934.

[GJJB05] James Gosling, Bill Joy, Guy L. Steele Jr., and Gilad Bracha. *The Java Language Specification*. Addison-Wesley Professional, Boston, 2005. ISBN 0-321-24678-0.

[Gra06] Elizabeth L. Gray. *A Trust-Based Reputation Management System*. PhD thesis, Trinity College, University of Dublin, Dublin, 2006.

[Gre11] Robert Gregory. *Psychological testing: History, principles and applications*. Pearson, 2011. ISBN 978-0-205-78214-7.

[Gro10] The PHP Group. DateInterval::format [online]. http://www.php.net/manual/en/dateinterval.format.php, 2014-02-11 [quoted 2014-05-10].

[Hen09] Anja Alide Jolijn Hendriks. *The construction of the Five-Factor Personality Inventory (FFPI)*. PhD thesis, Faculty of Behavioural and Social Sciences, University of Gronigen, Gronigen, 2009.

[JIB07] Audun Josang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. In *Decision Support Systems Volume 43, Issue 2*, Amsterdam, březen 2007. Elsevier Science Publishers B. V. Amsterdam.

[Koh02] Rudolf Kohoutek. *Základy užité psychologie*. Akademické nakladatelství CERM, Brno, 2002. ISBN 80-214-2203-3.

[Kol01] Kolektiv autorů. *Akademický slovník cizích slov*. Academia, Praha, 2001. ISBN 80-200-0607-9.

[Mac11] Tomáš Machulka. Metody analýzy socialní sítě pro dolování znalostí. Bachelor thesis, FIT BUT, Brno, 2011.

[Mar94] Stephen Paul Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, University of Stirling, Stirling, UK, 1994.

[Mas43] Abraham Maslow. A theory of human motivation. *Psychological Review*, page 50(4), 1943.

[MC85] Robert R. McCrae and Paul Costa. Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences*, (6):587–597, 1985.

[oT08] Brno University of Technology. Faculty of Information Technologies [online]. http://www.fit.vutbr.cz/.en, 2014-05-06 [quoted 2014-05-08].

[Oxf31] Social network definition [online]. http://www.oxforddictionaries.com/definition/english/social-network, 2013-06-10 [quoted 2013-12-31].

[Pau07] Daniela Pauknerová. *Psychologie pro ekonomy a manažery*. Grada Publishing, Praha, 2007. ISBN 80-247-1706-9.

[PHP31] PHP introduction [online]. http://www.w3schools.com/php/php_intro.asp, 2013-06-10 [quoted 2013-12-31].

[Rub64] Sergej Leonidovič Rubinštejn. *Základy obecné psychologie*. SPN, Praha, 1964. ISBN 16-906-67.

[Sam11] Jan Samek. *Důvěra a reputace v distribuovaných systémech*. PhD thesis, FIT BUT, Brno, 2011.

[Sch12] Bruce Schneier. *Liars and Outliers*. Wiley, 2012. ISBN 978-1118143308.

[Shi98] Lyndall Shick. *Understanding temperament*. Blue Cloud Books, Arizona, USA, 1998. ISBN 1-884734-32-4.

[Soc12] Social network [online]. http://en.wikipedia.org/wiki/Social_network, 2011-11-04 [quoted 2011-11-12].

[Svo05] Mojmír Svoboda. *Psychologická diagnostika dospělých*. Portál, Praha, 2005. ISBN 80-7367-050-X.

[The10] Theofrast [online]. http://www.denstoredanske.dk/Natur_og_milj%C3%B8/Botanik/Botanikere_og_plantegeografer/Theofrast, 2009-02-02 [quoted 2014-01-10].

[UMA31] UMAP 2013 [online]. http://www.dia.uniroma3.it/~umap2013/, 2013-06-10 [quoted 2013-12-31].

[ZTL+11] Lizi Zhang, Cheun Pin Tan, Siyi Li, Hui Fang, Pramodh Rai Yao Chen, Rohit Luthra, Wee Keong Ng, and Jie Zhang. The influence of interaction attributes on trust in virtual communities. In *International Conference on User Modeling, Adaptation and Personalization (UMAP)*. Nanyang Technological University, Singapore, 2011.

[ŠS13] Tomáš Švec and Jan Samek. Trust evaluation on facebook using multiple contexts. *CEUR Workshop Proceedings*, 2013(997):23–32, 2013.

[Šve11] Tomáš Švec. Vyhodnocení důvěry nebo reputace na základě analýzy sociální sítě. Bachelor thesis, FIT BUT, Brno, 2011.

[ŠČ04] Jitka Šimíčková Čížková. *Přehled sociální psychologie*. Univerzita Palackého v Olomouci, Olomouc, 2004. ISBN 80-244-0929-1.

[Řez98] Jaroslav Řezáč. *Sociální psychologie*. Paido, Brno, 1998. ISBN 80-85931-48-6.

[Říč07] Pavel Říčan. *Psychologie osobnosti*. Grada Publishing, Praha, 2007. ISBN 978-80-247-1174-4.

Appendix A

CD contents

The attached CD contains the following directory structure:

- **results**: contains the `.xlsx` file with collected results for reference
- **sources**: contains the sources and database settings needed to run the application
  - **db**: settings needed for creating the MySQL database
  - **application**: the sources in `.php` and `.css`
  - **correlation**: a small PHP application for finding correlation in given data
- **text**: the text of the thesis with full LaTeXsource code and also with figures in PDF

The working application registered under the Facebook ID 212701342136307 can be found running at http://www.stud.fit.vutbr.cz/~xsvect00/Facebook/.